



UNIVERSITAT DE
BARCELONA

Introducción del CAD (computer-aided detection) en un programa de detección precoz de cáncer de mama

Xavier Bargalló Castelló

ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (www.tdx.cat) i a través del Dipòsit Digital de la UB (diposit.ub.edu) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (www.tdx.cat) y a través del Repositorio Digital de la UB (diposit.ub.edu) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (www.tdx.cat) service and by the UB Digital Repository (diposit.ub.edu) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

UNIVERSIDAD DE BARCELONA
FACULTAD DE MEDICINA

TESIS DOCTORAL

**INTRODUCCIÓN DEL CAD (COMPUTER-AIDED DETECTION)
EN UN PROGRAMA
DE DETECCIÓN PRECOZ DE CÁNCER DE MAMA**

XAVIER BARGALLÓ CASTELLÓ

Directoras: Dra. Gorane Santamaría Ormaetxea
Dra. Carmen Ayuso Colella

Barcelona, 2016

Fácil es remediar cuando aún no han aparecido los síntomas

Tao Te Ching

A la Núria, la Laura i el Gerard

AGRADECIMIENTOS:

A mi madre, sin su ayuda no habría podido ser médico.

A la Dra. Gorane Santamaría y la Dra. Carmen Ayuso del servicio de radiología por sus comentarios y sugerencias para mejorar la tesis.

Al Dr. Martín Velasco fundador de la sección de radiología de mama del Hospital Clínic.

A todo el personal técnico y de enfermería de la sección de mamografía que de una forma u otra han colaborado para que esta tesis haya podido realizarse.

A todos mis compañeros radiólogos que han participado como lectores del programa de detección precoz.

A la Sociedad Española de Radiología Médica que colaboró con una beca SERAM para la realización del artículo 1 de esta tesis.

ÍNDICE

| | |
|---|-----------|
| ÍNDICE | 1 |
| I-INTRODUCCIÓN | 7 |
| 1.1-EPIDEMIOLOGÍA DEL CÁNCER DE MAMA | 9 |
| 1.1.1 Incidencia | 9 |
| 1.1.2 Factores de riesgo | 9 |
| 1.1.3 Pronóstico | 10 |
| 1.1.4 Prevención | 11 |
| 1.2- DETECCIÓN PRECOZ DEL CÁNCER DE MAMA | 12 |
| 1.2.1 Estudios poblacionales aleatorizados | 12 |
| 1.2.2 Controversia | 14 |
| 1.2.3 Efectos indeseables del cribado | 16 |
| 1.2.3.1 Sobrediagnóstico | 16 |
| 1.2.3.2 Falsos positivos | 19 |
| 1.2.3.3 Falsos negativos | 19 |
| 1.2.4 Posicionamiento del “Pla Director d’Oncologia de Catalunya” | 20 |
| 1.2.5 Funcionamiento del programa de cribado poblacional en Barcelona | 20 |
| 1.2.5.1 Antecedentes | 20 |
| 1.2.5.2 Proceso diagnóstico | 21 |
| 1.2.5.3 Indicadores de proceso | 23 |
| 1.3- PRUEBA DIAGNÓSTICA | 25 |
| 1.3.1 Mamografía | 25 |
| 1.3.1.1 Mamografía convencional | 26 |
| 1.3.1.2 Mamografía digital | 26 |
| 1.3.1.2.1 Detectores de fósforo fotoestimulable | 27 |
| 1.3.1.2.2 Sistemas con detectores integrados | 27 |
| 1.3.2 Rendimiento de la mamografía | 31 |
| 1.3.3 Doble lectura de la mamografía | 33 |
| 1.4- “COMPUTER-AIDED DETECTION” (CAD) | 37 |
| 1.4.1 Introducción histórica | 37 |
| 1.4.2 Funcionamiento de un sistema CAD | 38 |
| 1.4.2.1 Segmentación | 41 |
| 1.4.2.2 Realce de la imagen y normalización | 42 |

| | |
|--|-----------|
| 1.4.2.3 Extracción de la lesión y de sus características | 42 |
| 1.4.2.3.1 Detección de nódulos/asimetrías | 42 |
| 1.4.2.3.2 Detección de distorsiones de la arquitectura | 43 |
| 1.4.2.3.3 Detección de microcalcificaciones | 44 |
| 1.4.2.4 Análisis y clasificación de las características de la lesión | 45 |
| 1.4.2.5 Estudio comparativo de la mamografía | 45 |
| 1.4.3 Estado de la cuestión | 46 |
| 1.4.3.1 Estudios de demostración precoz de un beneficio potencial | 46 |
| 1.4.3.2 Estudios de valoración independiente del CAD | 47 |
| 1.4.3.3 Estudios de rendimiento “en laboratorio” | 48 |
| 1.4.3.4 Estudios de valoración clínica | 50 |
| 1.4.3.4.1 Lectura única versus lectura única + CAD | 50 |
| 1.4.3.4.2 Lectura única+ CAD versus doble lectura | 51 |
| II- JUSTIFICACIÓN | 55 |
| III- HIPÓTESIS Y OBJETIVOS | 59 |
| 3.1 Estudio 1 | 61 |
| 3.1.1 Hipótesis | 61 |
| 3.1.2 Objetivos | 61 |
| 3.2 Estudio 2 | 62 |
| 3.2.1 Hipótesis | 62 |
| 3.2.2 Objetivos | 62 |
| IV- PRESENTACIÓN DE LOS ARTÍCULOS PUBLICADOS | 63 |
| 4.1 Estudio 1 | 65 |
| “Role of computer-aided detection in very small screening detected invasive breast cancers” | 65 |
| Xavier Bargalló, Martín Velasco, Gorane Santamaría, Montse del Amo, Pedro Arguis y Sonia Sánchez Gómez. | |
| Journal of digital imaging. 2013; 26:572–577 | |
| IMPACT FACTOR: 1,2 QUARTIL 3 | |
| 4.1.1 Resumen de los resultados del estudio 1 | 65 |
| 4.1.1.1 Descripción de la muestra | 65 |
| 4.1.1.2 Sensibilidad y especificidad del CAD | 67 |
| 4.1.1.3 Factores que afectan el rendimiento del CAD | 68 |

| | |
|--|-----------|
| 4.1.1.4 CAD y doble lectura | 70 |
| 4.2 Estudio 2 | 71 |
| “Single reading with computer-aided detection performed by selected radiologists in a breast cancer screening program “ | 71 |
| Xavier Bargalló, Gorane Santamaría, Montse del Amo, Pedro Arguis, José Ríos, Jaume Grau, Marta Burrel, Enrique Cores y Martín Velasco. | |
| European journal of radiology. 2014; 83: 2019–2023 | |
| IMPACT FACTOR: 2,16 QUARTIL 2 | |
| 4.2.1 Resumen de los resultados del estudio 2 | 71 |
| 4.2.1.1 Rendimiento global del programa de cribado | 71 |
| 4.2.1.2 Rendimiento del CAD | 72 |
| 4.2.1.3 Estadificación e histología de los tumores detectados | 73 |
| 4.2.1.4 Rendimiento de los radiólogos | 74 |
| V- DISCUSIÓN | 75 |
| 5.1 Discusión de los resultados del estudio 1: | 77 |
| “Role of computer-aided detection in very small screening detected invasive breast cancers” | 77 |
| 5.2 Discusión de los resultados del estudio 2: | 82 |
| “Single reading with computer-aided detection performed by selected radiologists in a breast cancer screening program” | 82 |
| VI- CONCLUSIONES | 91 |
| VII- BIBLIOGRAFÍA | 95 |

I-INTRODUCCIÓN

1.1-EPIDEMIOLOGÍA DEL CÁNCER DE MAMA

1.1.1 Incidencia

El cáncer de mama es el cáncer más frecuentemente diagnosticado en la mujer excluyendo los cánceres de piel no melanoma (el carcinoma basocelular es el cáncer más frecuente en el ser humano). La mayoría de casos se presentan en mujeres postmenopáusicas con una mediana de edad de 61 años.

El número de casos nuevos anuales de cáncer de mama en Cataluña para los años 2003-07 fue de 3.907 lo que representó casi el 30 % del total de cánceres que se diagnosticaron en la mujer. La tasa de incidencia acumulada fue del 11,1 %, lo que supone que 1 de cada 9 mujeres padecerán cáncer de mama a lo largo de su vida ¹.

Según datos de la Sociedad Española de Oncología Médica (SEOM), el cáncer de mama en España durante 2012 fue el más incidente y el que causó más mortalidad en mujeres. Se diagnosticaron 25.215 nuevos cánceres de mama (29 % de todos los cánceres en mujeres) y 6.075 mujeres (15,5 % de los fallecimientos por cáncer) fallecieron por esta causa ².

1.1.2 Factores de riesgo

Factores de riesgo modificables son la obesidad, el consumo de alcohol, la terapia hormonal substitutiva y tabaquismo.

Factores de riesgo no modificables son el aumento de la densidad mamaria (relacionado con la cantidad de tejido fibroglandular respecto al adiposo medido en la mamografía), densidad ósea elevada, diabetes mellitus tipo II, hiperplasia glandular (ductal o lobulillar) atípica en una biopsia previa, carcinoma lobulillar in situ, radioterapia por cáncer en el tórax durante la

juventud, factores hormonales como una larga historia menstrual, uso reciente de anticonceptivos orales, no haber tenido hijos y tener el primer hijo después de los 30.

También es un factor de riesgo el tener familiares de primer grado afectados. Las mutaciones genéticas BRCA 1 y 2 que representan los cánceres de mama por susceptibilidad genética mejor estudiados, explican un 5-10 % de todos los cánceres de mama y hasta un 20 % de todos los cánceres con historia familiar.

La lactancia materna durante al menos un año, la actividad física moderada o intensa de forma regular, mantener un peso adecuado son factores que se asocian a una disminución del riesgo ³.

1.1.3 Pronóstico

El cáncer de mama es la segunda causa de muerte por cáncer en mujeres después del cáncer de pulmón. En Estados Unidos, las tasas de muerte por cáncer de mama han ido disminuyendo de forma sostenida desde 1989 con mayores descensos en mujeres más jóvenes. Así, entre 2007 y 2011, la tasa de muerte ha disminuido un 3,2 % por año en mujeres menores de 50 y un 1,8 % por año en mayores de 50 ⁴. La disminución de la mortalidad por cáncer de mama responde a mejoras relacionadas tanto con la detección precoz como con el tratamiento. La supervivencia media relativa del cáncer de mama tras cinco años es del 89,4 %. El estadio en el que se ha diagnosticado el cáncer influye en la supervivencia. Así, la supervivencia a los 5 años (2005-11) en estadios iniciales es de más del 98 %, si existe afectación de ganglios axilares es del 84,9 % mientras que en estadio IV, la supervivencia desciende al 25,9 % ⁵. De ahí la importancia del diagnóstico precoz, ya que se ha demostrado que el tamaño del tumor y la extensión ganglionar o a distancia tienen una relación inversamente proporcional con la supervivencia ^{6,7}. Diversos estudios han demostrado que el

sobrepeso afecta negativamente la supervivencia en la mujer postmenopáusica ^{8,9}. Del mismo modo, la actividad física también afecta positivamente a la supervivencia ¹⁰.

Según el registro de tumores de Tarragona ¹¹, en 1995-99 la supervivencia relativa a los 5 años era de 82,5% , que es superior a la europea (76,1 %) según datos de EURO CARE-3, e inferior a la americana (85,3%). En Cataluña, durante el periodo 2000-2004, la supervivencia fue del 82,8 % ¹.

1.1.4 Prevención

Podemos diferenciar entre prevención primaria y secundaria. La prevención primaria evita la adquisición de la enfermedad como ocurre por ejemplo en las campañas de vacunación o de intervención sobre riesgos ambientales o a través de la educación sanitaria. Su objetivo es prevenir la aparición de la enfermedad o daño en personas sanas. La prevención secundaria va dirigida a detectar la enfermedad en estadios precoces de la misma con el objetivo de que a través de medidas terapéuticas adecuadas se pueda curar o se impida su progresión.

En la actualidad, no es posible intervenir en la prevención primaria del cáncer de mama. Sin embargo, es factible realizar una detección precoz para detectar los tumores en estadios precoces con el objetivo de aumentar la supervivencia y además, favoreciendo tratamientos menos agresivos.

Para incluir una determinada enfermedad en un programa de cribado poblacional, ésta debe cumplir varios requisitos que ya fueron establecidos por Frame y Carlson en 1975 ¹²: a) debe ser una enfermedad que tenga un efecto significativo sobre la cantidad y calidad de la vida, b) deben existir métodos para tratarla de forma adecuada una vez que se ha detectado de forma precoz, c) debe tener un periodo asintomático inicial durante el cual su detección permita reducir la morbimortalidad así como un resultado terapéutico superior, d) la prueba

diagnóstica debe satisfacer criterios de coste-eficacia y debe ser bien aceptada por los pacientes, y e) la incidencia de la enfermedad debe ser suficiente para justificar el coste del programa de cribado, cuyo objetivo principal es reducir la mortalidad por cáncer.

El cáncer de mama cumple estas características ya que es una enfermedad grave y prevalente que se puede diagnosticar de forma precoz mediante la mamografía, que es una prueba asequible y los casos diagnosticados precozmente pueden ser tratados de forma más eficaz lo que conlleva un aumento de la supervivencia.

1.2- DETECCIÓN PRECOZ DEL CÁNCER DE MAMA

1.2.1 Estudios poblacionales aleatorizados

En las décadas de los 70 y 80 se realizaron varios ensayos comparativos aleatorizados que demostraron una reducción de la mortalidad por cáncer de mama en las mujeres invitadas a participar en programas poblacionales de cribado. El primer estudio lo inició el “Health Insurance Plan” (HIP) en 1963 en Nueva York ¹³ y en 1976 se inició el estudio sueco “Swedish two-county” ¹⁴. Ambos demostraron que los programas de cribado poblacional para la detección precoz disminuían la mortalidad por cáncer de mama. Posteriormente, se realizaron otros estudios suecos ^{15,16}, canadienses ¹⁷ y en Reino Unido ¹⁸.

A partir de estos estudios, la evidencia de que el cribado reduce las muertes por cáncer se ha establecido sólidamente y en la mayoría de países occidentales se han puesto en marcha programas poblacionales para la detección precoz del cáncer de mama. La figura 1 muestra el impacto en la reducción de la mortalidad de 9 estudios aleatorizados realizados entre 1963 y 1991 ¹⁹.

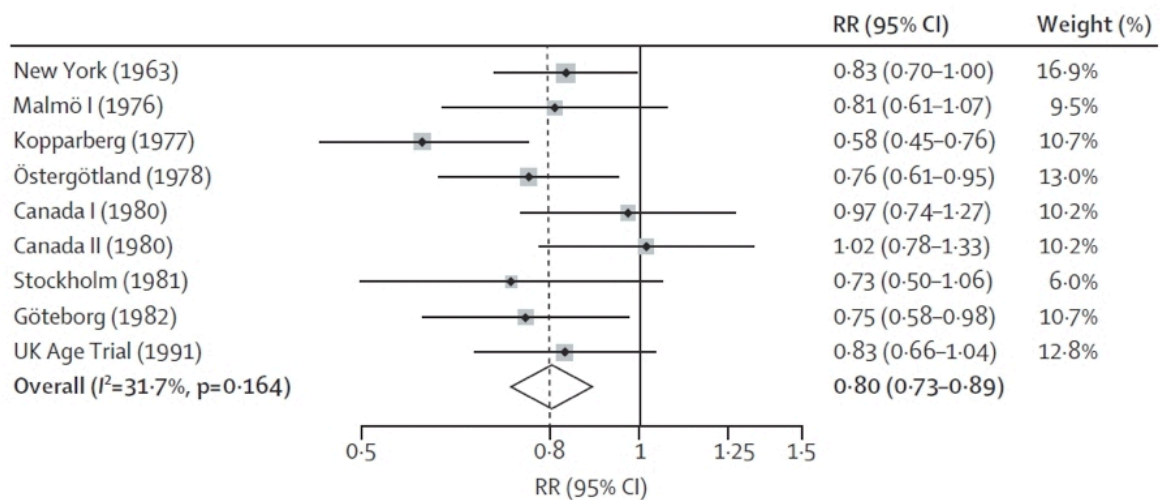


Figura 1. Riesgo Relativo (RR) de morir por cáncer de mama¹⁹

La mayoría de los estudios demuestran una disminución del riesgo relativo de morir por cáncer de mama de alrededor del 20 %, lo que indica que la participación en los programas de cribado tiene un efecto protector sobre la supervivencia. Sin embargo, el estudio de Canadá muestra unos resultados discordantes respecto al resto de estudios, por lo que merece especial atención. En este estudio se han reportado sesgos en la selección ya que se permitió participar a mujeres con cánceres palpables y en la aleatorización, que no fue ciega. Por otra parte, consultores externos al programa criticaron la calidad de las mamografías ya que hasta la mitad de las mamografías se juzgaron como de calidad baja o totalmente inaceptable. Además, no hubo una adecuada cooperación con los ginecólogos y cirujanos. De hecho, hasta casi un 25% de las localizaciones con aguja y biopsias que fueron recomendadas no llegaron a realizarse²⁰. Por tanto, quizás la conclusión del estudio de Canadá no debería ser que no hubo impacto en la mortalidad sino que la mamografía de baja calidad y la falta de coordinación e integración entre la parte radiológica del cribado y la parte clínica, reduce la efectividad del programa.

1.2.2 Controversia

En el año 2000 apareció una publicación en “The Lancet” por Olsen y Gotzsche en la que se revisaba la metodología de los ensayos que validaron el cribado poblacional de cáncer de mama y llegaron a la conclusión de que el cribado no solo no disminuía la mortalidad sino que se asociaba a tratamientos más agresivos^{21,22}. A raíz de este artículo se levantó una fuerte polémica en torno a los beneficios del cribado y el artículo ha sido analizado en profundidad. Por un lado, estos autores excluyeron del análisis ciertos ensayos que demostraban mayor beneficio del cribado y por otro lado, para medir el beneficio del cribado no usaron la disminución de mortalidad por cáncer de mama sino la reducción de mortalidad global, aduciendo que muchos errores se generaban al atribuir la causa de muerte. Aunque el razonamiento es correcto, el problema es que para que un ensayo detecte un impacto significativo de la disminución de la mortalidad por cáncer de mama en la mortalidad global debería incluir más de un millón de mujeres²³.

Por otra parte, la polémica suscitada llevó a analizar de nuevo de forma sistemática los ensayos aleatorizados incluidos teniendo en cuenta las críticas de Olsen y Gotzsche. En general, se ha concluido que el cribado de cáncer de mama con mamografía es eficaz para reducir la mortalidad en mujeres de más de 50 años²⁴. La nueva revisión de los estudios suecos concluyó que la mortalidad por cáncer de mama se reducía en un 21% (RR 0,79; IC 95%: 0,70-0,89) en el grupo de edad 40-74 años²⁵. La revisión de la “International Agency for Research on Cancer” (IARC) concluyó que en mujeres de 50-69 años la reducción era del 25 % (RR 0,75; IC 95%: 0,67-0,85)²⁶. Recientemente, la revisión de la “Canadian Task Force on Preventive Health Care” consideraba que se podía seguir recomendando el cribado periódico mediante mamografía en mujeres de 50-69 años, ya que se asociaba a una reducción de la

mortalidad del 21% (RR 0,79; IC 95%: 0,68-0,90) ²⁷ . La última revisión de la “US Preventive Services Task Force” obtuvo una reducción de mortalidad de 14 % (RR 0,86; IC 95%: 0,75-0,79) en mujeres de 50-59 años y una reducción de 32 % (RR 0,68; IC 95%: 0,54-0,87) para mujeres de 60-69 años ²⁸. Incluso evidencias más recientes que provienen de estudios observacionales, indican que en programas de cribado centralizados y organizados el beneficio es mayor del observado en los estudios aleatorizados poblacionales, pudiendo llegar a una reducción de la mortalidad del 50 % ^{23,29} .

Recientemente, la Organización Mundial de la Salud (OMS) ha publicado un informe que recomienda el cribado siempre que sea dentro de programas organizados y que se proporcione una adecuada información a la mujer ³⁰ .

Para estudiar el tema en profundidad, se reunió en Reino Unido un comité de expertos independientes que incluía expertos en epidemiología y estadística que no hubieran publicado en cáncer de mama para revisar la evidencia a favor y en contra del cribado. Este comité, el “Independent UK Panel on Breast Cancer Screening”, concluyó en un meta-análisis de 11 ensayos aleatorizados, que el riesgo relativo de morir por cáncer de mama en mujeres invitadas al cribado comparando con mujeres control era de 0,80 (IC 95%: 0,73-0,89), lo que significa una disminución en el riesgo relativo de muerte del 20%. Los autores del estudio afirmaban también que si se tuviera en cuenta la ventana ideal para hacer el seguimiento del efecto del cribado, es decir, cinco años después de iniciado el programa y hasta 10 años después de haberlo finalizado, el beneficio sería mayor ³¹.

También hay que tener en consideración dos factores que podrían ocasionar un sesgo contrario al beneficio del cribado como son el hecho que las mujeres invitadas y que rechazan participar en el programa se cuentan como cribadas y en segundo lugar, que las mujeres no invitadas pueden cribarse de forma oportunista y de este modo ser diagnosticadas de un

cáncer de mama en fase precoz. Por ejemplo, 10.000 de las 31.000 mujeres invitadas en el programa HIP de Nueva York no participaron (no se realizaron una mamografía) y fueron incluidas en el grupo cribado, diluyendo por tanto el beneficio que la mamografía aporta en este grupo. Por otro lado, hasta 26 % de las mujeres en el brazo control del estudio de Canadá se realizaron mamografías fuera del estudio y fueron consideradas no cribadas a pesar del beneficio que la mamografía les aportó ³².

En estudios observacionales se ha hecho el seguimiento de la causa de muerte durante 10 años demostrando que de 609 mujeres muertas por cáncer de mama en ese periodo, un 29 % habían participado en el programa de cribado (19 % cánceres detectados en el cribado y 10 % cánceres de intervalo), y un 71 % de las muertes ocurrieron en mujeres que no habían participado ³³.

Recientemente, el grupo EUROSCREEN ha llevado a cabo una revisión sistemática de la literatura, incluyendo estudios europeos de tendencia (n=17), estudios de incidencia de mortalidad (n=20) y estudios de caso control (n=8). En los estudios de incidencia de mortalidad la reducción fue del 25 % (RR 0,75; IC 95 %: 0,69-0,81) entre mujeres invitadas y del 38 % (RR 0,62; IC 95 %: 0,56-0,69) entre aquellas que realmente se habían hecho la mamografía. En los estudios de caso control la reducción fue de 31% (odds ratio [OR] 0.69; IC 95 %: 0.57-0.83) y 48 % (OR 0.52; IC 95 %: 0.42-0.65) respectivamente. Así pues, concluyen que la mortalidad ha disminuido entre un 25-31 % en mujeres invitadas y entre un 38-48 % en mujeres que realmente se han realizado la mamografía ³⁴.

1.2.3 Efectos indeseables del cribado

1.2.3.1 Sobrediagnóstico

Se define como la detección de un cáncer de mama confirmado histológicamente pero que

nunca se habría diagnosticado clínicamente durante la vida de la mujer. No obstante, hay que tener en cuenta que es una teoría puramente epidemiológica que solo puede ser valorada a nivel poblacional ya que actualmente es imposible identificar aquel cáncer que eventualmente podría no progresar o incluso regresar en base a criterios clínicos, radiológicos o anatomopatológicos. El sistema ideal para calcular el sobrediagnóstico es a partir de estudios aleatorizados que, en principio, no fueron diseñados para tal fin, sino para valorar la reducción de mortalidad. Para ajustar la anticipación en el diagnóstico que se asocia con el cribado deberían ser incluidos únicamente datos de estudios aleatorizados que no ofrecieron cribado al grupo de control al final de la intervención y que reportaron la incidencia de cáncer de mama hasta 5 años después³⁵. Según esta premisa, Moss *et al.* estimaron que en los ensayos canadienses el sobrediagnóstico fue del 14% y del 11 % a los 13 años de seguimiento en mujeres más jóvenes y mayores respectivamente³⁴. Analizando el ensayo de Malmo, Zackrisson *et al.* estimaron un 10 % de sobrediagnóstico a los 15 años de seguimiento³⁶. Según el “Independent United Kingdom Panel on Breast Cancer Screening” (IUKPBCS) solo hay tres estudios (Malmo I y los dos canadienses) que cumplan la premisa anterior aunque en alguno de éstos el seguimiento posterior fue solo de 6 años. Los resultados son de una estimación del sobrediagnóstico del 10,7 % para las invitadas al cribado, aunque si tenemos en cuenta las realmente cribadas sería del 20 %³¹.

La mayoría de los estudios observacionales para estimar sobrediagnóstico han analizado tendencias temporales o diferencias geográficas en la incidencia de cáncer de mama. Estos métodos llevan aparejados diversos sesgos, lo que explica que los resultados de las estimaciones sean ampliamente diferentes, oscilando entre un 4 %³⁷ y un 52 %³⁸. Así pues, dentro de los estudios observacionales son preferibles los estudios basados en cohortes. Bleyer y Welch estimaron que el 31 % de todos los cánceres diagnosticados en USA en 2008

no habrían progresado y por tanto no se hubieran diagnosticado durante la vida de la mujer ³⁹. No obstante, Helvie *et al.* estudiando los mismos datos valoraron que no se había tenido en cuenta el incremento en la incidencia global del cáncer en las tres últimas décadas y obtuvieron unas cifras del 1,3 % ⁴⁰. Respecto a los estudios que usaron cohortes, Olsen *et al.* obtuvieron una cifra de 4,8 % ⁴¹ y Puliti *et al.*, que siguieron una cohorte de 51.096 mujeres de 50-69 durante 15 años, estimaron un sobrediagnóstico del 10% ⁴².

Dado que el diagnóstico de carcinoma in situ ha aumentado en relación a los programas de cribado con una incidencia de alrededor del 20%, algunos autores opinan que estos casos podrían explicar una buena parte de los porcentajes de sobrediagnóstico. Al respecto, comentaré los siguientes estudios. Page y Dupont, en una extensa revisión de biopsias realizadas hasta 1982, detectaron 28 mujeres que habían sido diagnosticadas e intervenidas quirúrgicamente de enfermedades benignas pero que en realidad padecían un carcinoma ductal in situ de bajo grado. Estas mujeres fueron seguidas a lo largo del tiempo y se objetivó que 11 de ellas desarrollaron un carcinoma invasivo en el mismo cuadrante del que fueron intervenidas. Siete de ellas lo desarrollaron en los primeros 10 años de seguimiento y las 4 restantes más tarde, lo que podría también incidir en el hecho de que se necesiten largos tiempos de seguimiento para ajustar la anticipación en el diagnóstico que causan los programas de cribado. Cinco de estas 11 mujeres desarrollaron metástasis ^{43,44}. En el ensayo de United Kingdom, Australia y Nueva Zelanda sobre el tratamiento del carcinoma in situ se demostró que en el grupo que fue tratado únicamente con exéresis quirúrgica amplia recidivaron un 19 % de casos, la mitad de ellos en forma de cáncer invasivo ⁴⁵.

En resumen, dado que los estudios aleatorizados poblacionales se llevaron a cabo mucho tiempo atrás, antes de los actuales avances técnicos en el diagnóstico y tratamiento del cáncer de mama, sería recomendable que los programas de cribado abordaran el problema del

sobrediagnóstico mediante un estudio de cohortes para una mejor evaluación. Este estudio debería ser llevado a cabo a través de un grupo multidisciplinar para valorar la historia natural del cáncer de mama, que probablemente cambiará en los próximos años a través del estudio genético de los tumores, y valorar, a la luz de este nuevo conocimiento, qué clase de tumores se diagnostican en los programas de cribado y cuál es su historia natural más probable ⁴⁶.

1.2.3.2 Falsos positivos

Se definen como aquellos casos en que se realizan pruebas radiológicas adicionales para estudiar una alteración visible en la mamografía de cribado y en los que finalmente se descarta la existencia de un cáncer. El riesgo acumulado de la ocurrencia de un falso positivo en una mujer de 50-69 años que participa en el programa de manera bianual durante los 20 años durante los que será invitada a participar en el programa es de un 19,7 %. Si consideramos aquellos casos en los que para excluir malignidad se han tenido que realizar procedimientos invasivos, el riesgo sería del 2,9 % ⁴⁷. No obstante, este porcentaje varía mucho según los programas implicados y la experiencia de los radiólogos.

1.2.3.3 Falsos negativos

El cáncer de intervalo es aquel cáncer detectado después de una mamografía de cribado normal y antes de la próxima mamografía de cribado (es decir, antes de dos años). Pueden ser cánceres que se pasaron por alto en la anterior mamografía o cánceres más agresivos que se manifiestan en el período entre cribados. Se ha estimado que alrededor del 30 % son falsos intervalos ⁴⁸; es decir, falsos negativos por error de lectura del radiólogo y el resto son verdaderos intervalos, es decir, con mamografía previa negativa. De todos los cánceres detectados en mujeres que participan en un programa de cribado se considera que entre 16-28 % son cánceres de intervalo ⁴⁹.

1.2.4 Posicionamiento del “Pla Director d’Oncologia de Catalunya”

Para finalizar este apartado presentaré las conclusiones de una reciente revisión de la comisión asesora del cribado del “Pla Director d’Oncologia”. Esta comisión estima que en términos de impacto absoluto, por cada mil mujeres de 50-69 años cribadas cada 2 años durante 20 años y seguidas hasta los 79 años, se evitarán 8 muertes por cáncer de mama; 200 mujeres tendrán un falso positivo y 30 de ellas recibirán un procedimiento intervencionista; finalmente, se podrían sobrediagnosticar 4 cánceres de mama. Así, pues, esta comisión recomienda que el cribado poblacional para la detección precoz del cáncer de mama debe continuar, mejorando la información en cuanto a riesgos y beneficios que reciben tanto los profesionales sanitarios como las mujeres que participan en el programa de forma que puedan realmente tomar una decisión informada sobre su participación⁵⁰.

1.2.5 Funcionamiento del programa de cribado poblacional en Barcelona

El cáncer de mama cumple los requisitos comentados antes (Frame y Carlson¹²) para poder ser incluida en un programa de cribado poblacional del cáncer de mama .

1.2.5.1 Antecedentes

El programa de detección precoz de cáncer de mama en la ciudad de Barcelona se puso en marcha en noviembre de 1995. El Consorcio Sanitario de Barcelona encargó al Hospital del Mar la puesta en marcha de la fase piloto del actual programa que se inició en Ciutat Vella y Sant Martí. Durante 1998 se incorporó el Hospital del Vall d’Hebrón cribando el barrio del Carmelo al que posteriormente se sumaron los distritos de Nou Barris y Horta-Guinardó. A continuación, el Hospital de la Esperanza convocó a las mujeres de Gracia y Sarrià-Sant Gervasi. Durante el año 2001 se incorporó el Hospital de Sant Pau que cribó a las mujeres de

la derecha del Eixample y Sant Andreu y durante 2002 se sumó el Hospital Clínic cribando la izquierda del Eixample, Sants-Montjuïc y Les Corts. Durante 2004, el Guinardó pasó al Hospital de Sant Pau.

La población diana del programa son todas las mujeres residentes en la ciudad de Barcelona de 50 a 69 años. Este grupo recibe una carta del hospital proveedor del servicio invitándolas a realizarse una mamografía cada 2 años que es el tiempo que tarda el programa en citar a toda la población diana de su zona. Cada uno de estos intervalos de 2 años constituye una ronda de cribado.

1.2.5.2 Proceso diagnóstico

La detección del cáncer de mama en el programa de cribado puede ser considerado un proceso en dos fases. En primer lugar la paciente se realiza una mamografía que consiste en la obtención de dos proyecciones de cada mama: craneocaudal y mediolateral oblicua. En los modernos mamógrafos digitales directos (también llamados de campo completo) las imágenes quedan almacenadas en un sistema "Picture Archiving and Communication System" (PACS) que es un sistema de archivo y transmisión de imágenes médicas, para ser visualizadas en una estación de trabajo dedicada a patología mamaria que va acompañada de dos monitores de alta resolución (5 millones de píxeles). El radiólogo revisa las 4 imágenes y las compara con estudios previos siempre que sea posible. El resultado puede ser a) negativo o benigno, por lo que se realizará el próximo control a los 2 años; b) probablemente benigno, generalmente reservado para controlar microcalcificaciones (se estima que en esta categoría la probabilidad de malignidad es menor al 2 %), por lo que se realizará el próximo control en un año y c) insuficiente para diagnóstico o sospechoso de malignidad en cuyo caso pasará al circuito asistencial de la unidad de mama donde la mujer entra en un segundo nivel de

estudio en el que se realizarán diversas pruebas que pueden consistir en proyecciones mamográficas adicionales o focalizadas, ecografía e incluso biopsia con aguja gruesa, todas ellas dirigidas a descartar o confirmar malignidad.

El informe radiológico sigue las directrices marcadas por la Clasificación BI-RADS (“Breast Imaging Reporting and Data System”) del Colegio Americano de Radiología ⁵¹:

-BI-RADS 1: Negativo

-BI-RADS 2: Hallazgo benigno con toda seguridad.

-BI-RADS 3: Hallazgo probablemente benigno

-BI-RADS 0: Hallazgo insuficiente para diagnóstico. Requiere de una prueba diagnóstica adicional para realizar un diagnóstico de seguridad. Esta prueba puede ser simplemente una proyección mamográfica adicional o con frecuencia una ecografía y en algunas ocasiones biopsia mediante aguja gruesa guiada por ecografía o estereotaxia.

-BI-RADS 4: Hallazgo sospechoso de malignidad que será estudiado mediante biopsia con aguja gruesa. Este grupo a su vez se subdivide en tres grupos: a, b y c según la probabilidad creciente estimada de malignidad.

-BI-RADS 5: Hallazgo altamente sospechoso de malignidad (probabilidad > 90%). La biopsia es obligada.

En Barcelona, el proceso de lectura sigue el protocolo de doble lectura con arbitraje de los casos discordantes, es decir, aquellos en los que un radiólogo recita para estudiar un hallazgo determinado y el otro radiólogo no, ya sea porque no lo ha visto o bien porque lo valora como benigno.

1.2.5.3 Indicadores de proceso

Definiciones de los principales indicadores de proceso que se utilizan en el programa de cribado poblacional:

-Población diana: mujeres entre 50 y 69 años residentes en la ciudad de Barcelona.

-Población invitada: mujeres que una vez realizadas las exclusiones que determina el programa, son invitadas a participar.

-Población participante: mujeres que acuden a la cita y se realizan una mamografía como consecuencia de la invitación por parte del programa.

-Motivos de exclusión: defunción, diagnóstico previo de cáncer de mama, errores del padrón (mujeres de las que se sabe que ya no viven en la ciudad), enfermedad terminal y-o mental grave que impidan la realización de la mamografía y mujeres que se dan de baja voluntariamente del programa.

-Mujeres incluidas en primer cribado: mujeres que participan por primera vez en el programa.

a) Primera invitación: participan la primera vez que son invitadas al programa.

b) Invitación previa: son mujeres que habían sido convocadas en rondas anteriores y no habían participado. Sin embargo, en la ronda actual deciden participar por primera vez.

-Mujeres incluidas en segundo cribado o sucesivos: mujeres que participan en el programa y ya lo habían hecho en alguna ronda anterior.

a) Regular: han participado en todas las rondas anteriores.

b) Irregular: no han participado en todas las rondas anteriores.

-Cobertura: mujeres que se realizan la mamografía tanto dentro como fuera del programa. Se obtiene sumando las mujeres participantes con aquellas que se realizan la mamografía fuera del programa ya sea en el sistema público o privado.

-Recitación: mujeres en las que el resultado de la mamografía es positivo (BI-RADS 0-4-5) y que por tanto son llamadas y citadas para acudir al circuito asistencial y determinar si realmente tienen una lesión o no. En el caso de que tengan una lesión se realizará una biopsia para determinar la naturaleza de la misma.

-Pruebas invasivas: aquellos casos en los que para determinar la naturaleza de la lesión es necesario recurrir a una biopsia que puede ser una punción con aguja fina o una biopsia con aguja gruesa o mediante aspiración por vacío o incluso una biopsia quirúrgica.

-Controles avanzados: aquellos casos que son tributarios de un control antes de los dos años, ya sea como resultado directo de la mamografía o después de haber sido estudiadas en el circuito asistencial.

Tal como muestra la tabla 1, durante el 2013 la población diana del programa en Barcelona fue de 100.833 mujeres de las que 96.001 fueron invitadas a participar. De éstas, finalmente participaron 50.735 mujeres, es decir, un 52,8 % de las invitadas siendo el grupo de 65-69 años el que tuvo una mayor participación. La cobertura global (teniendo en cuenta las mujeres que se realizan la mamografía fuera del programa) fue del 73,6 %. Un 6,6 % de las mujeres fueron recitadas para la realización de pruebas adicionales y en un 1,5 % fue necesario practicar una prueba invasiva. Al final del proceso diagnóstico se diagnosticaron 202 cánceres lo que representó 4 mujeres diagnosticadas de cáncer por cada mil participantes⁵².

Tabla 1. Principales indicadores de proceso durante 2013 en la ciudad de Barcelona estratificados por tipo de cribado ⁵²

| | 1er cribado Primera invitación | 1er cribado Invitación previa | 2º cribado regular | 2º cribado irregular | TOTAL 2013 |
|-------------------------------|--------------------------------------|-------------------------------------|-----------------------|-------------------------|---------------|
| Invitadas | 14.549 | 27.946 | 44.747 | 8.759 | 96.001 |
| Participantes | 6.173 | 2.870 | 38.721 | 2.971 | 50.735 |
| % Participación | 42,4 | 10,3 | 86,5 | 33,9 | 52,8 |
| Mujeres con mamografía | 2.843 | 11.939 | 2.216 | 2.919 | 19.917 |
| % Cobertura | 62,0 | 53,0 | 91,5 | 67,2 | 73,6 |
| Recitadas | 948 | 410 | 1.725 | 257 | 3.340 |
| % Recitación | 15,4 | 14,3 | 4,5 | 8,7 | 6,6 |
| Pruebas invasivas | 199 | 77 | 415 | 54 | 745 |
| % Invasivas | 3,2 | 2,7 | 1,1 | 1,8 | 1,5 |
| Controles avanzados | 213 | 106 | 330 | 50 | 699 |
| % Controles avanzados | 3,5 | 3,7 | 0,9 | 1,7 | 1,4 |
| Cánceres detectados | 25 | 16 | 147 | 14 | 202 |
| Tasa x 1000 | 4,0 | 5,6 | 3,8 | 4,7 | 4,0 |

1.3- PRUEBA DIAGNÓSTICA

1.3.1 Mamografía

Es una prueba radiológica para el estudio de la mama que utiliza radiación ionizante y que consiste en realizar dos radiografías de cada mama en proyección cráneo-caudal (el haz de rayos X atraviesa la mama de arriba abajo) y mediolateral-oblicua (el haz de rayos X atraviesa la mama de forma lateral-oblicua). Es necesario realizar una moderada compresión de la mama para limitar al máximo la superposición de estructuras glandulares. La dosis glandular media de una mamografía analógica es de 4,7 miligrays (mGy) y de una digital de 3,7 mGy. La

radiación efectiva sobre la mama se ha estimado entre 0,4 y 0,7 milisieverts (mSv). Para ponerlo en perspectiva, la radiación anual de fuentes naturales está entre 2 y 3 mSv y una tomografía axial computarizada abdominal sin contraste implica 10 mSv^{53,54}.

1.3.1.1 Mamografía convencional

Es el clásico sistema de cartulina-película cuya principal ventaja es su alta resolución espacial (hasta 14 pares de líneas/mm) que posibilita la detección de las microcalcificaciones y de las finas espiculaciones asociadas a ciertos tipos de cáncer. Su principal desventaja es su estrecho rango dinámico que reduce la posibilidad de detectar diferencias sutiles de densidad especialmente en mamas densas. Otra limitación importante es el ruido estructural asociado al grano de la película y a la pantalla de refuerzo. Además, presenta también inconvenientes relacionados con el flujo de trabajo y el manejo de pacientes. Este sistema requiere un proceso de revelado de la película que enlentece el proceso de obtención de la mamografía y que requiere de un estricto control para que no merme la calidad.

1.3.1.2 Mamografía digital

En la actualidad los avances en los detectores ofrecen una respuesta lineal en un amplio rango (1000:1) comparado con el sistema cartulina/película (40:1)⁵⁵. Otra ventaja es que la adquisición se realiza utilizando detectores de rayos X altamente eficientes y con un nivel de ruido muy bajo. Además, la imagen se almacena digitalmente y se procesa utilizando filtros diseñados específicamente para optimizar la visualización de las estructuras y detalles de interés diagnóstico. En el proceso de lectura, el radiólogo puede ajustar el brillo, contraste y aumento de la imagen. No obstante, los detectores digitales tienen una resolución espacial peor debido a los problemas que conlleva hacer detectores con píxeles de pequeño tamaño que cubran el campo completo de la mama. Las imágenes son visualizadas en monitores de

alta resolución (5 millones de píxeles). El almacén digital de la imagen permite acceder a mamografías previas con facilidad para comparar con ellas la mamografía actual.

Aunque la mamografía convencional tiene una mayor resolución que la digital, varios estudios han demostrado que la resolución espacial del sistema digital no es el factor más limitante en la detección de microcalcificaciones sino que valores bajos de la relación señal-ruido o de la relación señal-contraste pueden ser más cruciales en la detección de estructuras de muy pequeño tamaño, por lo que la mamografía digital puede proporcionar un mayor rendimiento pese a su menor resolución espacial ^{56,57,58}.

1.3.1.2.1 Detectores de fósforo fotoestimulable

Conocidos como “computed radiography” (CR). Su gran ventaja es que pueden utilizarse con equipos mamográficos convencionales ya que consisten en placas de fósforos fotoestimulables que se introducen dentro de un chasis similar al del sistema cartulina/película que posteriormente requieren equipos lectores de CR para generar la imagen mamográfica digital. Este tipo de detectores tienen una eficiencia peor que los detectores integrados y suelen requerir de mayor dosis de radiación para obtener una calidad aceptable ⁵⁹. Además, en un gran estudio de cohortes dentro de un programa organizado de cribado poblacional se ha publicado que la tasa de detección de cáncer es menor para los sistemas CR ⁶⁰.

1.3.1.2.2 Sistemas con detectores integrados

Conocidos como “direct radiography” (DR) y también como “full-field digital mammography” (FFDM). Estos sistemas son rigurosamente digitales y tienen integrado el equipo de rayos X con el detector.

a) Producen una imagen inmediata, sin procesos intermedios de revelado. Hacen desaparecer los “chasis” con lo que permiten construir un entorno limpio puramente digital con tiempos muertos menores ya que el técnico que realiza la mamografía obtiene en su pantalla una imagen de la mama de forma casi inmediata lo que posibilita incrementar el rendimiento de la sala.

b) Reducen las dosis a los pacientes o al menos no las incrementan ya que la eficiencia de los detectores es mayor.

c) Producen imágenes de calidad muy apreciable, de forma mucho más estable. Además, su resolución de contraste es muy superior.

Existen diversas tecnologías en el mercado sin que sea posible todavía dilucidar si alguna de ellas es superior a la otra en cuanto a detección de cáncer.

1.3.1.2.2.1 DETECTORES DE SELENIO AMORFO

Se denominan de conversión directa porque el detector es un fotoconductor de selenio amorfo que convierte directamente los fotones de rayos X en pares electrón-hueco.

Estos detectores tienen mayor eficiencia cuántica en la detección de los rayos X en el intervalo de las energías utilizadas en mamografía (20-30 keV) comparado con los detectores de yoduro de cesio que explicamos en el siguiente apartado. Este método minimiza la dispersión lateral, garantizando una imagen de gran nitidez y mayor resolución espacial. La carga es leída por una matriz activa en contacto directo con la superficie inferior de selenio.

La limitación clásica de estos detectores es una cierta remanencia de la imagen adquirida asociada a la persistencia de cargas eléctricas residuales lo que exigirá aplicar técnicas de borrado de la imagen.

1.3.1.2.2.2 DETECTORES DE YODURO DE CESIO SOBRE SILICIO AMORFO

Se denominan de conversión indirecta. En este caso se utiliza una capa de material centelleante que es el yoduro de cesio que emite luz al absorber los fotones de rayos X. El centelleador se deposita sobre una capa de material fotoconductor de silicio amorfo cuya misión es transformar la luz producida en la lámina fluorescente en cargas eléctricas (figura 2). En la trayectoria del fotón de luz entre centelleador y lámina de silicio se producen fenómenos de difusión lateral que tienden a disminuir la resolución espacial del conjunto. Para corregir esto se utilizan centelleadores estructurados en los que los cristales tienen forma de agujas de pocas micras de diámetro que se apilan lateralmente actuando como guías de la luz. Con esta estructura el espesor puede ser mayor mejorando la eficiencia de detección cuántica (DQE) y minimizando las pérdidas en resolución espacial. El buen acoplamiento entre la capa de yoduro de cesio y la matriz de silicio amorfo permite obtener un sistema con una DQE muy elevada y, por tanto, una eficiencia muy buena. A ello contribuye también el tamaño de pixel (100 micras de lado), que si bien limita la resolución espacial a 5 pares de líneas/mm, también permite mejorar el contraste al reducir el ruido cuántico en cada pixel de la imagen, debido al aumento de superficie de cada pixel ⁶¹.

Este sistema es el que utiliza el mamógrafo digital Senograph 2000D[®] de General Electric con el que se han realizado todas las mamografías de los estudios que forman esta tesis. El detector tiene unas dimensiones de 19 x 23 cm con un tamaño de pixel de 100 micras de lado y una profundidad de bits de 14.

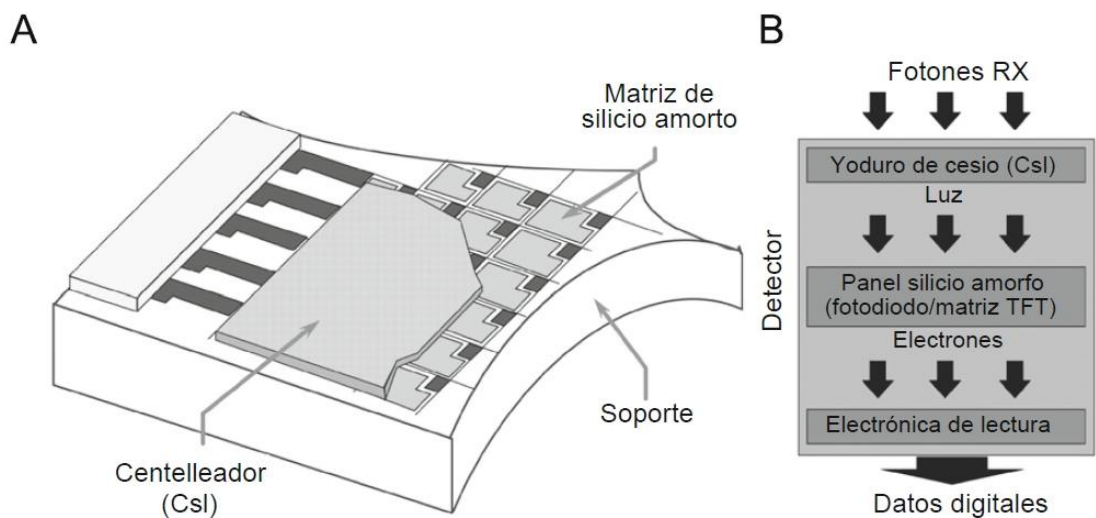


Figura 2. A. Estructura de un panel plano de silicio amorfo. B. Proceso esquemático de adquisición de imágenes en un panel plano de conversión indirecta ⁶¹.

1.3.1.2.2.3 DETECTORES DE RECuento DE FOTONES

Los detectores descritos hasta ahora se basan en la recolección de la carga producida por los rayos X, actúan como detectores integradores de energía, es decir, la imagen resultante está relacionada con la energía total absorbida por el material detector. En cambio, los detectores de recuento se basan en el recuento individual (fotón a fotón) de los fotones de rayos X incidentes. El detector que absorbe los fotones es de silicio cristalino que está dividido por placas de aluminio con una separación de 50 micras que es lo que determina el tamaño del pixel (50 x 50 micras). Una de las ventajas de estos detectores es la eliminación del ruido asociado con los procesos de conversión de los fotones de rayos X lo que puede asociarse a una disminución de la radiación necesaria.

La FDA aprobó el primer mamógrafo digital de tipo DR para su uso en pacientes en el año 2000 y a partir de ese momento se abrió un gran debate en cuanto a su valoración frente al

sistema convencional aunque en la actualidad está ampliamente reconocida la calidad de la mamografía digital. Se han publicado tres grandes estudios prospectivos comparando la mamografía tradicional con la digital en la detección del cáncer de mama. En dos de ellos, conocidos como Oslo I y II no se encontraron diferencias estadísticamente significativas en la detección de cáncer ^{62,63}. En el tercer estudio conocido como “Digital Mammography Imaging Screening Trial” (DMIST) se demostró además que en mujeres de menos de 50 años pre o perimenopáusicas y en mujeres con mamas densas, la mamografía digital tenía un rendimiento superior a la convencional ⁶⁴.

1.3.2 Rendimiento de la mamografía

La sensibilidad de la mamografía para detectar cáncer en pacientes sintomáticas oscila entre 81-96 %. La sensibilidad de la mamografía en pacientes asintomáticas (población diana de los programas de cribado) para detectar cáncer oscila entre el 70-90 % ⁶⁵. Esta variabilidad depende fundamentalmente de los siguientes factores:

- a) Densidad de la mama, que suele correlacionarse con la edad de la mujer. Así, cuanto más joven es la mujer, más densa es la mama por el predominio del componente fibroglandular y, paralelamente, menos sensibilidad de la mamografía. Por tanto, en mujeres de menos de 40 años la sensibilidad está en torno al 54-58% y en mujeres de más de 65 años, que suelen presentar mayor cantidad de tejido adiposo en la mama es del 81-94% ⁶⁶. Las mujeres con mamas extremadamente densas tienen un riesgo 6 veces mayor de ser diagnosticadas de cáncer de intervalo ⁶⁷. Este motivo junto con el hecho de que los tumores en mujeres jóvenes suelen ser más agresivos son los que explican que el cribado poblacional en mujeres de 40-50 años sea un tema controvertido ^{68,69,70}.

- b) Posicionamiento de la paciente. Con frecuencia tumores muy posteriores o muy mediales pueden quedar fuera del alcance de la mamografía si el posicionamiento no es muy cuidadoso. Además, con cierta frecuencia ocurre que una lesión situada en la periferia de la mama solo se ve en una proyección lo que dificulta la valoración del hallazgo y puede causar un falso negativo ⁷¹.
- c) Experiencia y nivel de especialización del radiólogo. Existe controversia sobre los factores que influyen en el rendimiento del lector de mamografía de cribado. Se ha investigado la influencia del número de mamografías leídas al año aunque uno de los factores más influyentes es que el radiólogo realice procedimientos de mamografía diagnóstica e intervencionista, lo que generalmente ocurre en radiólogos especialmente dedicados a la mama ^{72,73,74}.
- d) Tipo histológico del cáncer. El carcinoma lobulillar tiene cierta tendencia a crecer respetando la arquitectura mamaria y causa menor reacción desmoplásica, lo que hace que sea más difícil de detectar.
- e) Manifestación radiológica. Si el cáncer de mama muestra calcificaciones es más fácil de detectar. Por el contrario, si produce una distorsión de la arquitectura puede pasar fácilmente desapercibido ya que es el patrón radiológico que requiere más experiencia para ser detectado ⁷⁵.

Por todo ello, tal y como ya recomendó Wolfe en 1966, una mamografía negativa nunca debería demorar una biopsia clínicamente indicada. Esta premisa sigue totalmente vigente en la actualidad pese a los avances que ha supuesto la mamografía digital.

Varios estudios han demostrado que entre un 25-30 % de los cánceres de intervalo ya eran visibles y merecedores de recitación en la mamografía previa ^{76,77}. Los cánceres no detectados pueden atribuirse a errores de detección o errores de caracterización. Los primeros se deben fundamentalmente a fatiga del radiólogo, al tipo de manifestación radiológica o a la densidad de la mama que hace que la lesión se superponga a estructuras fibroglandulares que pueden disimularla. Birdwell *et al.* encontraron que hasta en un 30% de casos falsos negativos podían verse calcificaciones de forma retrospectiva ⁷⁸. Los errores de caracterización consisten en que el radiólogo realmente ha visualizado el hallazgo, pero ya sea por falta de experiencia, por la sutileza del mismo o porque no muestra signos de sospecha radiológica, lo valora como benigno y no recita a la paciente para completar su estudio.

La especificidad de la mamografía de cribado varía considerablemente según se trate de una primera ronda o rondas sucesivas y también dependiendo del país. Así, la especificidad en Estados Unidos para primera ronda y segunda o sucesivas es de 83,2 % y 91,6 % respectivamente. El programa danés aporta unas cifras de 96,6 % y 98,8 % respectivamente ⁷⁹. En nuestro programa, la especificidad para segunda ronda o sucesivas en el periodo 2010-12 estimando una tasa de cáncer de intervalo de 1 por mil es de 93 %.

1.3.3 Doble lectura de la mamografía

El proceso de lectura de la mamografía puede ser muy complicado teniendo en cuenta que el objeto de la búsqueda (cáncer de mama) solo está presente en 4-6 de cada mil mamografías que se informan. En este sentido, varios estudios han demostrado una moderada variabilidad afectando tanto a la detección como a la interpretación de los hallazgos radiológicos. ^{80,81}.

Por todo ello, se introdujo la idea de la doble lectura de la mamografía de cribado como mejor sistema para minimizar errores. A partir de este punto, diferentes estudios han demostrado sin excepción un aumento en la detección de cáncer que oscila entre un 5 y un 15 % ^{82,83,84,85} y que, por lo general, se asocia de forma inversamente proporcional a un aumento de la recitación (falsos positivos) que puede llegar a tres puntos porcentuales ⁸⁴; de forma que los estudios sobre precisión diagnóstica en forma de “área bajo la curva” muestran que la doble lectura no aumenta la precisión ⁸⁵ ya que para aumentar la detección genera un número alto de falsos positivos lo que disminuye la especificidad. Para contener este aumento de falsos positivos que genera la doble lectura, se ha demostrado que el arbitraje por un tercer radiólogo funciona eficazmente ⁸⁶.

Parece evidente, pues, que la doble lectura aumenta la sensibilidad, es decir ayuda a detectar más cánceres. No obstante hay que tener en cuenta varios factores.

- a) El problema de la escasez de radiólogos, sobre todo con experiencia suficiente para leer mamografía de cribado, está bastante extendido y está documentado tanto en Inglaterra ^{87,88} como en Estados Unidos ⁸⁹. Sobre cómo solucionar este tema hay varias aproximaciones. En Estados Unidos prácticamente no existe la doble lectura. En países como Inglaterra, Países Bajos y Canadá se ha intentado solucionar el problema de la escasez de radiólogos formando técnicos en radiología para emparejarlos con radiólogos ^{88,90,91,92}.
- b) De cualquier modo, aunque no se recurra a técnicos, es un hecho que existen pocos radiólogos con dedicación total o parcial a la imagen diagnóstica de mama por lo que es usual emparejar a radiólogos más expertos con otros con menos experiencia tal como indican los rangos de sensibilidades de lectura publicados ^{87,93}. Por otra parte,

muchos programas utilizan un sistema de doble lectura que no es independiente; es decir, el segundo lector ve las marcas o tiene los resultados del primero y lee sobre lo leído de forma que no es posible saber cuál es su sensibilidad real (sabemos lo que aporta el segundo al primero pero no lo que aporta el primero al segundo) lo que hace más complicado analizar los resultados.

- c) Para contener la recitación se ha recurrido en general al arbitraje que cumple este cometido de forma eficaz pero con el inconveniente de que a su vez también genera falsos negativos. Es decir, el árbitro tiene que despreciar algunos hallazgos vistos por uno de los lectores que eventualmente resultan ser un cáncer. Este hecho ha sido puesto de manifiesto en varios estudios que evidenciaron que entre un 3 y un 4 % de los informes negativos del árbitro fueron cáncer ^{94,95}.

Para ilustrar todo ello con cifras, en el programa holandés la sensibilidad interlectores para la detección de cáncer de mama osciló entre un 51,5 y un 75 % con unos valores predictivos positivos de 29,5 hasta 45,4 %. Las variaciones en la tasa de detección de cáncer oscilaron entre 3,4 y 5 % ⁸⁷. Estos datos son similares a los del estudio americano de Gur *et al.* que observaron un tasa de detección de cáncer entre 2,6 y 5,4 % ⁹⁶ y también con los de nuestro propio programa en el que la tasa de detección de cáncer entre diferentes lectores durante el periodo histórico (2004-10) osciló entre 2,73 y 7,08 % con una tasa de recitación entre 3,11 y 10,88 %.

El problema expuesto en el punto b, tiene como consecuencia que hay un lector experto que aporta muchos cánceres a la lectura del inexperto y no al revés. Por ejemplo, en un estudio del programa holandés dos tercios de aquellos cánceres detectados por el segundo lector lo fueron por dos de los 8 lectores, lo que indica claramente la importancia de la experiencia del

radiólogo en los resultados de la doble lectura. Además, encontraron diferencias significativas en la sensibilidad de las parejas de lectores dependiendo de si incluían o no un radiólogo experto en mama. Así pues, recomienda como deseable formar parejas que incluyan radiólogos dedicados a la mama aunque reconoce que esto es poco realista dada la menor proporción de radiólogos con dedicación completa a la mama⁹⁷. En un estudio del programa británico, dos de los 5 radiólogos que formaban los emparejamientos mostraban un rendimiento muy por debajo del resto en cuanto proporción de cánceres no vistos por ellos y si por su pareja. Tres de los radiólogos fallaron entre 4 y 6 cánceres mientras que los otros dos fallaron 16 uno y 45 el otro. Así pues, es fácil imaginar que las parejas estaban claramente compuestas por uno experto y otro no, siendo la direccionalidad de la aportación clara⁹³. Otros estudios corroboran este hecho y relacionan el rendimiento del radiólogo en mamografía de cribado con el número de exploraciones diagnósticas de mama y de biopsias de mama que ha realizado y no encuentran relación con el volumen anual de mamografías leídas ni con los años que llevan haciéndolo. Es decir, rinden más aquellos radiólogos que participan en la parte asistencial del diagnóstico mamario que son los que se dedican a la radiología asistencial de mama de forma total o parcial^{74,98}.

En resumen, podríamos decir que la doble lectura aumenta la detección de cáncer de mama pero que el efecto de la misma probablemente ha sido sobreestimado ya que se suelen emparejar radiólogos con diferente nivel de experiencia y precisión diagnóstica con lo que en general, el que rinde más aporta al que rinde menos. Por otra parte, la doble lectura aumenta la recitación de forma marcada en algunos casos. Para mantener la recitación en unos márgenes aceptables se ha recurrido en muchos programas al arbitraje. Este por su parte también está sujeto a tener falsos negativos.

1.4- “COMPUTER-AIDED DETECTION” (CAD)

1.4.1 Introducción histórica

Ya en la década de los 70 se empezó a contemplar la posibilidad de que la tecnología computacional, que apenas empezaba a mostrar sus posibilidades, pudiera usarse en el diagnóstico radiológico ⁹⁹. En esa época se asumió que los computadores podrían reemplazar a los radiólogos, pero los resultados no fueron buenos debido a que los ordenadores eran muy primitivos, no estaban disponibles técnicas avanzadas para el procesamiento de imágenes y las imágenes digitales prácticamente no existían. Más adelante, los avances en la ciencia informática y un cambio en la manera de abordar el problema permitieron obtener los primeros resultados. Simplemente, la idea de un ordenador que diagnostica se cambió por la más sutil de un ordenador que detecta. En este caso, el ordenador no pretende substituir al radiólogo sino que le ayuda en la detección de lesiones.

Los primeros trabajos para intentar diseñar un sistema informático de ayuda al diagnóstico radiológico comenzaron a mediados de los años 80 cuando un equipo de físicos y radiólogos de los laboratorios Kurt Rossmann del departamento de radiología de la Universidad de Chicago unieron sus esfuerzos para tal fin. En aquel momento, se puso énfasis en la detección de nódulos en la radiografía de tórax y en la detección de microcalcificaciones en la mamografía ¹⁰⁰, de modo que el papel del CAD es el de ayudar al radiólogo y no substituirlo ^{101,102}. De este equipo de investigadores surgió en 1987 la primera patente sobre CAD, que se ha convertido con el tiempo en la más citada en el campo de esta tecnología y sobre la cual se han desarrollado los diferentes CAD comerciales.

El primer prototipo de CAD fue desarrollado en la misma Universidad de Chicago en noviembre de 1993 ¹⁰². El prototipo incluía un scanner de láser para digitalizar las

mamografías y una estación de trabajo con una librería de discos ópticos y grandes monitores de alta resolución que se presentó en Chicago, durante el congreso americano de radiología. Se invitó a 118 radiólogos de mama a proporcionar casos para probar el sistema CAD. Los resultados de este test de validación fueron prometedores¹⁰³.

1.4.2 Funcionamiento de un sistema CAD

El cáncer de mama se manifiesta en la mamografía de diversas formas siendo cinco los patrones más frecuentes: a) la agrupación de microcalcificaciones, b) nódulos, c) distorsiones de la arquitectura, d) asimetrías focales y e) una combinación de los anteriores. El CAD solamente discrimina entre calcificaciones, que marca como rectángulos (Figura 3) y el resto de lesiones, esto es, nódulos, asimetrías o distorsiones que marca como óvalos (Figura 4). Ello es así por la propia dificultad del CAD en diferenciar estos patrones entre sí y por el hecho que muchas lesiones muestran varios patrones radiológicos de forma combinada siendo frecuente ver una asimetría o nódulo con componente distorsionante.

El CAD hace meramente de asistente en la lectura, su función queda muy bien expresada como “segunda opinión” pero no vinculante. Es decir, el radiólogo interpreta la mamografía sin ayuda del CAD y una vez ha tomado una decisión, vuelve a revisar la mamografía pero esta vez con las marcas CAD y valora la posibilidad de que éstas realmente indiquen patología. A partir de aquí el radiólogo toma una decisión sobre la conducta a seguir. De este modo, el CAD alerta al radiólogo de posibles anomalías pero, en definitiva, es el radiólogo el que finalmente decide si acepta la sugerencia del CAD y recita a la paciente para continuar el estudio de la imagen señalada o bien la descarta porque juzga que se trata de una superposición de estructuras o de una lesión de naturaleza benigna.

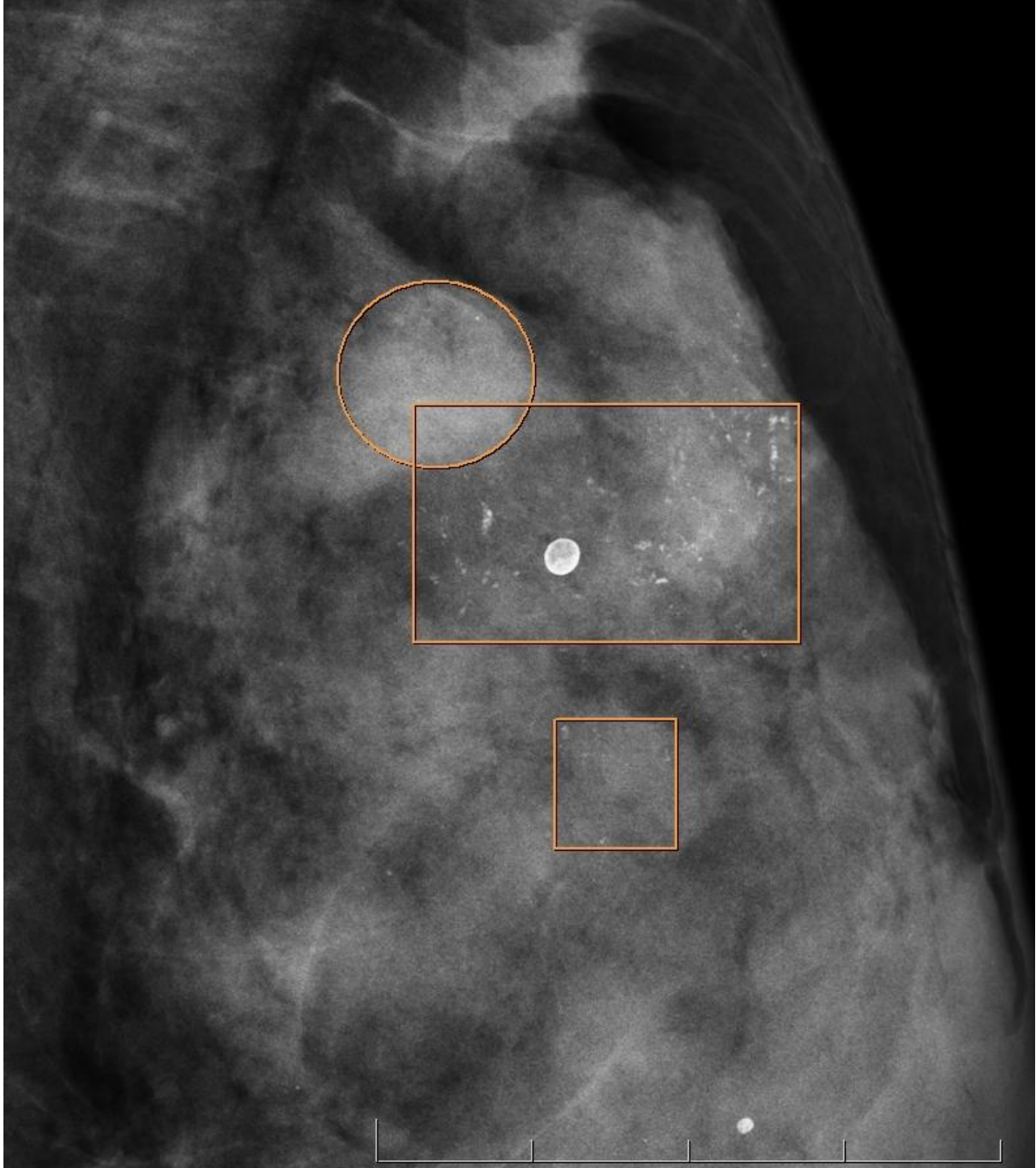


Figura 3. Proyección mediolateral oblicua izquierda magnificada. Agrupación de microcalcificaciones pleomorfas que en el estudio histológico correspondían a un carcinoma ductal in situ de 29 mm. La tinción inmunohistoquímica demostró un foco de microinfiltración de 1 mm. El CAD marca correctamente las calcificaciones (rectángulos) y coloca una marca falsa positiva (círculo).

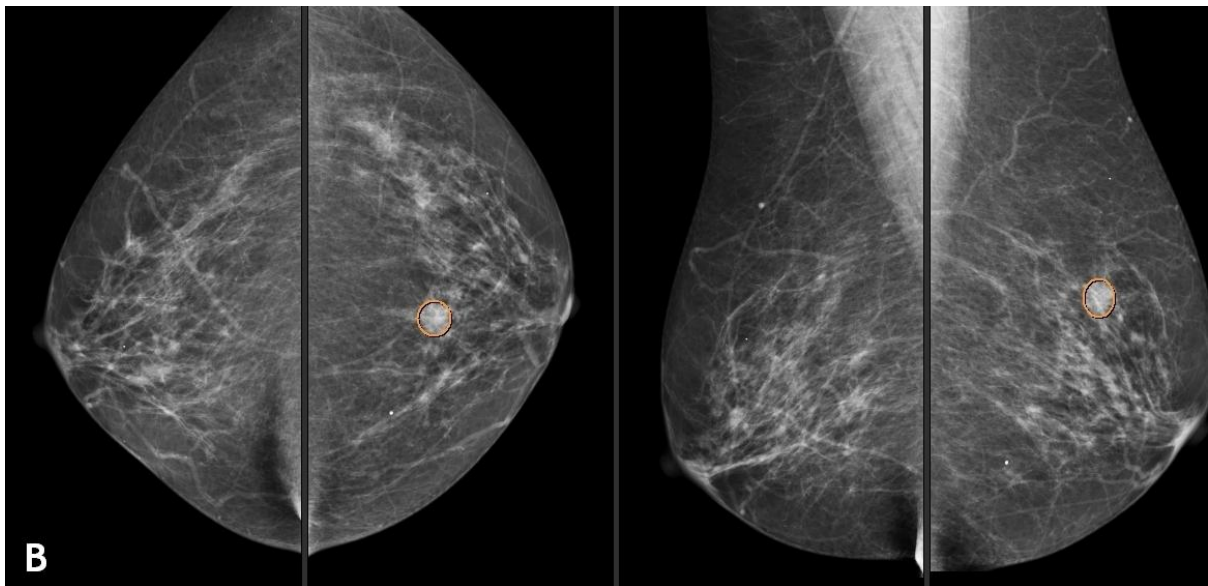
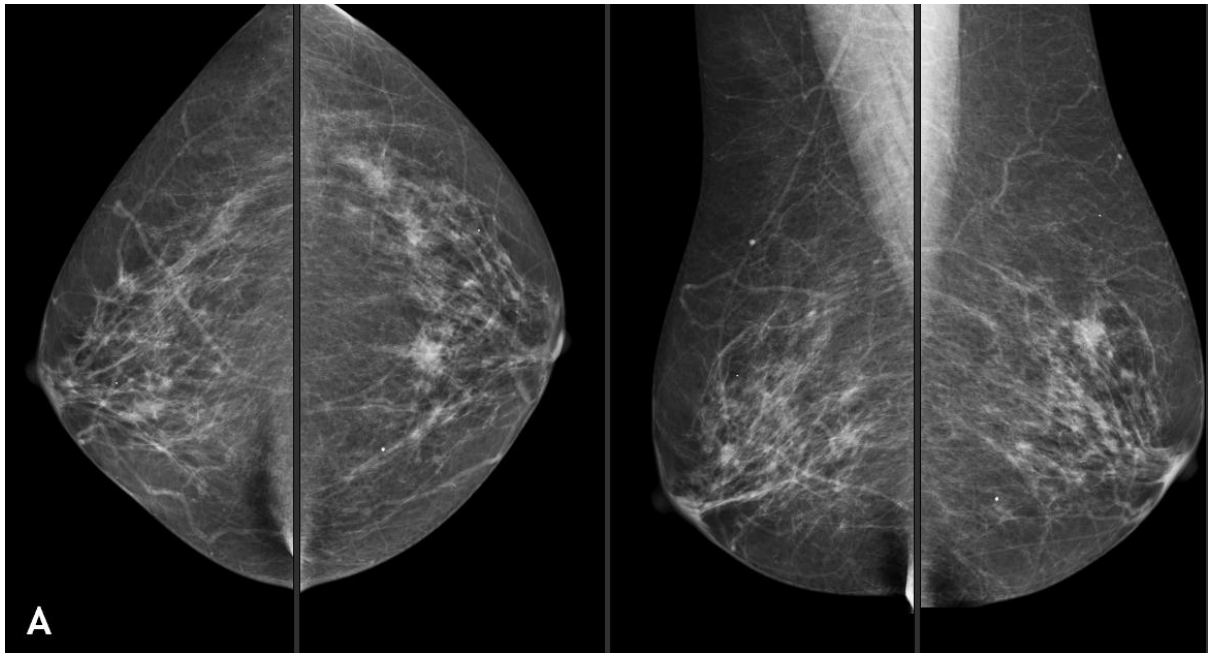


Figura 4. Mamografía bilateral. A. Imagen nodular mal delimitada en UCS de mama izquierda. B. El CAD marca correctamente el tumor en las dos proyecciones. El estudio histológico demostró un carcinoma ductal infiltrante de 11 mm

En la figura 5 se puede observar resumidamente cómo funciona un sistema CAD ¹⁰⁴.

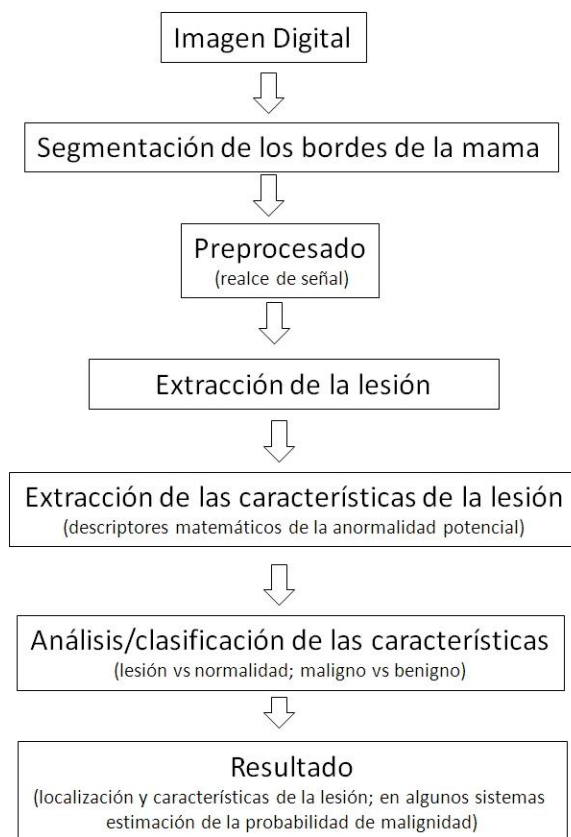


Figura 5. Etapas del trabajo con la imagen mamográfica realizado por un sistema CAD

1.4.2.1 Segmentación

Segmentar una imagen significa identificar las diferentes partes que la constituyen. Así, el primer paso es identificar y separar diferentes regiones de la mamografía, que típicamente son el fondo, tejido fibroglandular mamario, músculo pectoral, la línea de la piel y el pezón. Para segmentar se utilizan métodos de “umbralización” (del inglés “thresholding”) que definiendo un valor umbral de intensidad de pixel, convierten una imagen en escala de grises en una nueva con solo dos niveles, de manera que los objetos queden separados del

fondo.

1.4.2.2 Realce de la imagen y normalización

Si disponemos de mamografías en placa radiográfica, se precisa su transformación a formato digital mediante un scanner de radiografías. Posteriormente, se normalizan ajustando a una resolución y rango de intensidad del pixel determinados para poder ser analizadas de forma efectiva por el sistema CAD.

La mamografía digital directa (DR) nos proporciona imágenes nativas digitales en las que disponemos del total de los datos brutos proporcionados por el detector lo que implica una mayor cantidad de información y más homogénea.

1.4.2.3 Extracción de la lesión y de sus características

1.4.2.3.1 Detección de nódulos/asimetrías

En la terminología BI-RADS un nódulo se define como una lesión ocupante de espacio que muestra un borde convexo visible en las dos proyecciones; si solo se ve en una proyección se denomina asimetría.

Los métodos para detectar este tipo de lesiones inicialmente buscan regiones donde podría haber una lesión ocupante de espacio para posteriores análisis más precisos. Uno de los más usados es el “método de crecimiento de regiones” que en el caso más simple utiliza una única propiedad del pixel que suele ser la intensidad de gris. Según la intensidad fijada se selecciona un punto de máxima intensidad (pixel-semilla) que se va expandiendo a los pixeles vecinos con características similares. No obstante, estos algoritmos tienen problemas para extraer lesiones que están parcialmente obscurecidas por tejido fibroglandular normal y también en detectar espiculación¹⁰⁵.

Para intentar superar estos problemas se han desarrollado “métodos basados en modelos”. Aquí el objetivo primario no sería identificar los píxeles del nódulo correctamente sino generar únicamente una señal que localice el nódulo y posteriormente se abordará el problema de segmentar o delimitar los bordes de la lesión. Estos métodos se basan en el uso de una base de datos de casos de cáncer que el sistema CAD utiliza para comparar con las mamografías que analiza. Esta comparación se realiza usando algoritmos de inteligencia artificial que lo que hacen es buscar características que sugieran malignidad (irregularidad de contornos, espiculación, etc). Con este enfoque se intenta evitar que el CAD marque cualquier área densa sin tener en cuenta sus características.

1.4.2.3.2 Detección de distorsiones de la arquitectura

Las distorsiones incluyen los siguientes patrones: a) espículas finas o gruesas radiando de un centro con frecuencia virtual (signo de la estrella) y b) retracción o irregularidad focal del contorno glandular afectando la zona periférica de la mama en la unión entre grasa y parénquima (signo del borde o del hachazo). El patrón distorsionante es el tercer patrón más frecuente de cáncer de mama no palpable de modo que hasta un 15 % de los cánceres invasivos de menos de 1 cm se manifiestan como distorsiones¹⁰⁶. Con frecuencia los falsos negativos de lectura mamográfica se asocian con este patrón radiológico que suele ser sutil y difícil de detectar tanto para el radiólogo como para el CAD^{75,77}. En la figura 6 se observa una distorsión no marcada por el sistema CAD.

Las espículas están causadas por la reacción desmoplásica que causa el cáncer y son un importante signo de sospecha tanto en nódulos espiculados como en distorsiones por lo que un sistema CAD debería ser capaz de detectarlas. Desde el punto de vista de la imagen son estructuras lineales que parten de la lesión en una distribución radial y que pueden

confundirse con otras estructuras lineales de la mama como son vasos y conductos mamarios. Los algoritmos más utilizados para su detección se basan en el uso de funciones matemáticas que permiten detectar en la imagen elementos orientados en una determinada dirección y con un determinado grosor, o que son capaces de extraer líneas que pasan cerca del centro de la masa. No obstante, solo consiguen altas sensibilidades a expensas de muchas marcas falsas positivas ¹⁰⁷.

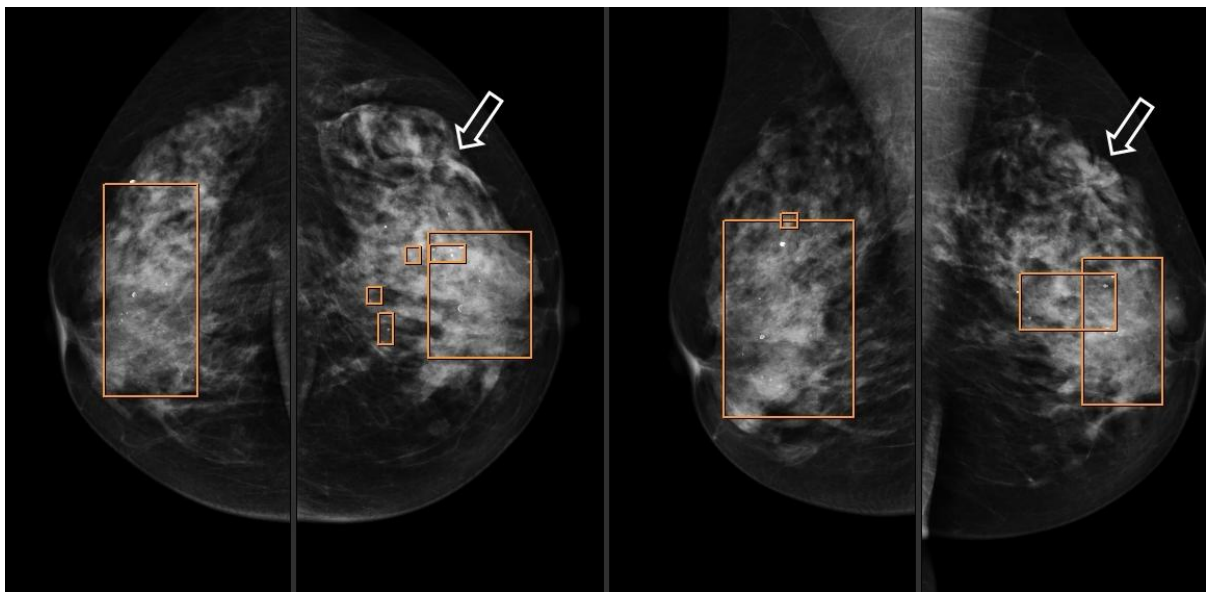


Figura 6. Mamografía bilateral. Distorsión de la arquitectura en el cuadrante supero-externo de mama izquierda visible en las dos proyecciones (flechas). El estudio histológico demostró un carcinoma ductal infiltrante de 18 mm. El CAD pone numerosas marcas sobre las calcificaciones displásicas bilaterales (rectángulos) que pueden despistar al radiólogo de la verdadera patología que no está marcada.

1.4.2.3.3 Detección de microcalcificaciones

Las microcalcificaciones son diminutos depósitos de calcio que causan pequeños puntos hiperdensos en la mamografía. Las microcalcificaciones agrupadas (más de 5 calcificaciones en 1 cm²) pueden ser un indicador de cáncer. Entre 30-50 % de cáncer de mama se manifiesta de forma única o combinada con calcificaciones.

Las microcalcificaciones son más densas que el tejido circundante, lo que en términos de

análisis computacional de la imagen se traduce en píxeles más brillantes que el tejido fibroglandular en las que están inmersas. El sistema CAD aplica varios filtros para potenciar el contraste y aislar la calcificación respecto del tejido mamario circundante ¹⁰⁸.

1.4.2.4 Análisis y clasificación de las características de la lesión

Una vez aplicados estos sistemas de extracción, se obtienen una serie de regiones candidatas que suelen estar entre 5 y 25 por proyección mamográfica y que deben pasar un proceso de “selección” para determinar aquellas que ofrecen una probabilidad de lesión maligna más alta y que serán las que el CAD marcará y de esta forma poder limitar el número de falsos positivos ¹⁰⁹. Para ello, el CAD aplica estrategias dirigidas a determinar la morfología, la textura y los contornos de la lesión. Además, el sistema CAD dispone de una base de datos de neoplasias que utiliza también para la clasificación final de la lesión aplicando sistemas de inteligencia artificial ¹¹⁰.

1.4.2.5 Estudio comparativo de la mamografía

Para el radiólogo es crucial el poder leer la mamografía de forma simétrica, es decir, estudia de forma comparativa las dos mamas, enfrentando las dos proyecciones craneocaudales por una parte y las dos oblicuas por otra, para valorar eventuales asimetrías. Para el CAD esta confrontación es muy difícil de realizar de forma que la mayoría de sistemas CAD actuales se basan en el análisis de cada proyección de forma independiente. Se están investigando métodos de geometría proyectiva para predecir dónde debería situarse una lesión en la otra proyección. Wei *et al.* han desarrollado un sistema CAD que interpreta la información proveniente de otras proyecciones utilizando un árbol de decisión, de forma similar a como lo hace el radiólogo. Con este sistema han reportado un aumento en la sensibilidad de 58 a 76 % con un rango de falsos positivos de 0,5 marcas por imagen ¹¹¹.

Las versiones más modernas de CAD no solo participan en la asistencia a la detección, sino que han implementado también un sistema que caracteriza las imágenes proporcionando una puntuación al hallazgo marcado. Es decir, asignan una probabilidad de malignidad lo que, de algún modo, contribuye a la caracterización de la imagen. Este enfoque retoma la idea original de diagnóstico por ordenador y se ha denominado CADx (Computer-aided Diagnosis) en frente del CAD de uso clínico actual que es meramente una herramienta de detección y se denomina CADe (Computer-aided Detection).

1.4.3 Estado de la cuestión

A continuación, se revisan los estudios que se han realizado para validar y aplicar la tecnología CAD.

1.4.3.1 Estudios de demostración precoz de un beneficio potencial

En las primeras fases de desarrollo de una nueva tecnología es importante demostrar su potencial para valorar si merece la pena continuar con los esfuerzos de desarrollo y mejora del producto. Un método particularmente útil en este caso fue la valoración retrospectiva de las mamografías de los casos falsos negativos. De este modo, se valoró si el CAD era capaz de marcar casos que representaban un cáncer incipiente y que el radiólogo no consideró para su recitación en ese momento, pero que retrospectivamente se juzgó que deberían haber sido recitados para ampliar estudio.

Warren Burhenne *et al.* revisaron las mamografías previas de una serie de 427 cánceres de mama y comprobaron que una tercera parte de casos mostraban hallazgos en la mamografía previa que de forma retrospectiva fueron considerados indeterminados o

sospechosos y, por tanto, merecedores de una recitación para continuar el estudio. Estudiaron el papel del CAD en estos casos y comprobaron que el CAD podía detectar hasta un 77 % de estos hallazgos. También vieron que la sensibilidad de los radiólogos hubiese mejorado de un 78,8 % hasta un 95,2 % (incremento del 21%) si se hubiese usado CAD ¹¹². Brem *et al.* en un estudio de similares características pero usando otra marca comercial de CAD reportaron que el CAD detectaba hasta un 65 % de estos cánceres que ya mostraban signos radiológicos en la mamografía previa. También objetivaron un incremento en la sensibilidad del radiólogo del 21 % ¹¹³.

Claramente, estos estudios mostraban potencial pero no un beneficio tangible en un escenario clínico real ya que para ayudar al radiólogo a detectar cáncer deben darse dos premisas. La primera es que el CAD marque la lesión como se ha demostrado que ocurre en un porcentaje alto de casos y la segunda es que el radiólogo responda positivamente a la sugerencia del CAD.

1.4.3.2 Estudios de valoración independiente del CAD

Son estudios retrospectivos en los que se estudia el papel del CAD en una serie de casos de cánceres de mama y adolecen, por tanto, del mismo problema que los anteriores ¹¹⁴.

Ellis *et al.* encontraron una sensibilidad del CAD de 81,8 % y una especificidad de 39,2 % para carcinomas invasivos ≤ 15 mm. Los errores del CAD se dieron en lesiones iso o hipodensas (46%), inmersas en el tejido fibroglandular (29 %) o cercanas al pezón (13 %) y categorizadas como BI-RADS 4-A o 4-B ¹¹⁵. Kim *et al.*, utilizando una versión de CAD más moderna, reportaron una sensibilidad general de 96,2 %, una sensibilidad de 89,1 % para las lesiones tipo masa y una sensibilidad de 100% para calcificaciones ¹¹⁶.

Aunque la sensibilidad del CAD para detectar calcificaciones ha sido descrita como próxima al 100 %, existe cierta preocupación sobre si esto es aplicable en todos los tipos de calcificaciones. Las calcificaciones denominadas amorfas son muy tenues y de muy pequeño tamaño, por lo que no se puede precisar exactamente su forma. De todas las calcificaciones que se envían a biopsia con aguja gruesa, las amorfas representan un 34 %, son malignas en un 20 %, y corresponden a una lesión de alto riesgo en otro 20 %. Debido a que son tenues, con cierta frecuencia escapan al escrutinio rápido de una mamografía. Soo *et al.* encontraron que la sensibilidad del CAD en este tipo de calcificaciones era del 51 % (57 % en aquellas que resultaron ser malignas), lo que indica la dificultad del CAD para detectarlas debido a que su densidad es menor de la habitual ¹¹⁷.

Como se ha comentado anteriormente, las distorsiones son el patrón radiológico más difícil de detectar y requieren una mayor pericia y experiencia por parte del radiólogo. Baker *et al.* compararon dos sistemas diferentes de CAD y encontraron una sensibilidad que osciló entre el 49 % y el 33 % para detectar distorsiones, lo que ilustra las limitaciones de la informática para detectar estas lesiones que suelen mostrar un centro poco denso con finas espículas radiantes que están inmersas en el parénquima circundante quedando parcialmente enmascaradas ⁷⁵. En un estudio más reciente, Hupse *et al.* usando una tecnología CAD de desarrollo propio mostraron una sensibilidad del CAD similar a la del radiólogo para la detección de patrones sin calcificaciones aunque con una especificidad menor ¹¹⁸.

1.4.3.3 Estudios de rendimiento “en laboratorio”

Son experimentos conocidos como multilector-multicaso en los que en condiciones controladas se da un paquete de casos con un mayor número de cánceres que en la rutina normal a un número variado de radiólogos y se comprueba su rendimiento sin y con CAD.

Chan *et al.* realizaron uno de los primeros estudios de este tipo en el que 60 mamografías (la mitad contenían microcalcificaciones) fueron estudiadas por 15 radiólogos. Los resultados se evaluaron mediante curvas ROC que demostraron una mejora en el rendimiento de los radiólogos usando CAD ¹¹⁹. Otros estudios similares han mostrado también aumento de la sensibilidad en la detección de lesiones ^{120,121}.

Los estudios de rendimiento en laboratorio son experimentos rigurosos y bien controlados y pueden producir valoraciones claras y nada ambiguas de cualquier nueva tecnología. No obstante, su punto fuerte es a la vez su punto débil; es decir, dado que se han realizado en condiciones “de laboratorio”, surgen dudas de cómo se hubieran comportado los mismos lectores en un contexto clínico real. El hecho de que un paquete de casos tenga una proporción mucho más alta de neoplasias (entre 30-50 % de la muestra habitualmente) respecto a lo que ocurre en la práctica rutinaria en que es muy baja (4-6 por mil) también incide en el comportamiento de los radiólogos.

Aunque todos los estudios comentados en los apartados anteriores muestran unos resultados prometedores, deben ser obligatoriamente refrendados en la práctica clínica para valorar la interacción real que se produce entre el CAD y el radiólogo ¹²². El radiólogo inmerso en la rutina diaria en que tiene que valorar muchas mamografías que son mayoritariamente normales puede tender a desestimar algunas marcas del CAD. Por otra parte, aunque está claro que el CAD marca lesiones que resultan ser un cáncer, hasta un 40% de los casos falsos positivos para cáncer en mamografía son errores de caracterización o de decisión ¹²³.

1.4.3.4.1 Lectura única versus lectura única + CAD

Freer and Ulissey, en un estudio prospectivo, evaluaron el efecto del CAD en la mamografía de cribado de 12.897 mujeres. Usaron un diseño de lectura secuencial en el que primero leían la mamografía y después evaluaban las marcas del CAD sobre la misma. Reportaron un aumento en la tasa de detección de cáncer de 3,2 a 3,8 %, lo que supone un incremento del 19,5 %, y un aumento en la tasa de recitación del 6,5 al 7,7 % ¹²⁴. Otro estudio prospectivo evaluó el efecto del CAD en la mamografía de cribado de 21.349 mujeres ¹²⁵. En este estudio, el CAD marcó 8 cánceres que el lector no detectó y, de éstos, 5 se manifestaron como grupos de calcificaciones. El estudio concluyó que el CAD aumentaba el número de cánceres detectados en un 7,62 % y el número de recitaciones en un 9,5 %. Birdwell *et al.* evaluaron de forma prospectiva el efecto del CAD en 8.682 mamografías de cribado hechas y leídas en un hospital universitario aunque sólo dos de los siete radiólogos que participaron estaban especializados en imagen de mama. Comunicaron un incremento en la tasa de detección de cáncer del 7,4% (de 27 a 29 cánceres) ¹²⁶.

El metanálisis llevado a cabo por Taylor y Potts (Figura 7) demostró que el CAD no tenía un efecto significativo en la tasa de detección de cáncer (IC 95%: 0,96-1,13) (Figura 7) y aumentaba la tasa de recitación (IC 95%: 1,09-1,12) ¹²⁷. De modo similar, Noble *et al.*, en un metanálisis llevado a cabo posteriormente usando diferentes sistemas de valoración y criterios de inclusión más estrictos que los de Taylor y Potts, concluyeron en el mismo sentido que si bien se detectaban más cánceres con el CAD, no era una diferencia significativa. Según los datos que se desprenden de su estudio, por cada 10.000 mujeres cribadas, se detectarían 5 neoplasias, habría 600 recitaciones y 40 biopsias ¹²⁸.

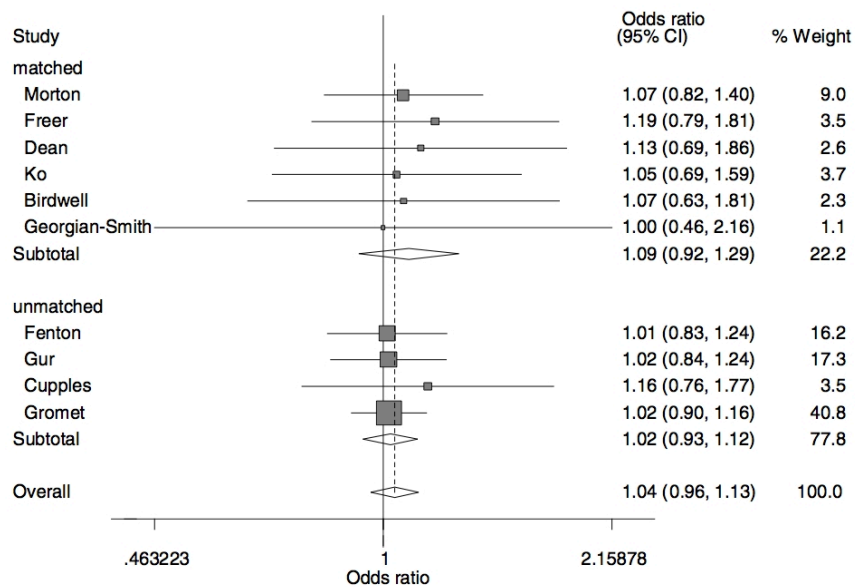


Figura 7. Gráfico obtenido del metanálisis de Taylor y Potts. En ninguno de los estudios revisados se observó una “odds ratio” significativa cuando se leía con CAD con respecto a lecturas sin CAD ¹²⁷

Fenton *et al.*, en un estudio multicéntrico, comunicaron que con el CAD se producía un aumento de sensibilidad en la detección de cáncer de mama de 80,4 a 84 % sin significación estadística ($P=0,32$) que se acompañaba de un aumento en las recitaciones y biopsias. La especificidad disminuía del 90,2 al 87,2 % ($P<0,001$), el valor predictivo positivo disminuía del 4,1 al 3,2 % ($P=0,01$) y la tasa de biopsia aumentaba en un 19,7 % ($P<0,001$) ¹²⁹. Este artículo ha sido muy criticado, fundamentalmente en cuanto a la heterogeneidad de las instalaciones de los centros participantes en el estudio y la experiencia dispar de los radiólogos participantes.

1.4.3.4.2 Lectura única+ CAD versus doble lectura

Este tipo de estudios son clínicamente más relevantes para valorar si el CAD puede realmente substituir a un lector. Los podemos clasificar en a) estudios transversales, cuando se realizan comparaciones sobre las mismas mamografías. Se controla mejor la variable a

estudiar, que es la influencia del CAD en la lectura, pero es difícil incluir un gran número de pacientes y b) estudios longitudinales, que permiten incluir un mayor número de pacientes aunque variables ajenas al estudio, difíciles de controlar, pueden afectar el resultado.

ESTUDIOS TRANSVERSALES

Un estudio prospectivo inglés comparó la lectura única + CAD con la doble lectura en mamografías de cribado de 6.111 mujeres. Doce cánceres no se vieron en la lectura única y, de éstos, nueve fueron correctamente marcados por el CAD. No obstante, 7 de las nueve marcas fueron despreciadas por el lector. Así pues, el CAD aumentó la sensibilidad de la lectura única en solo un 1,3 % y la doble lectura en un 8,2 % ¹³⁰. Otro estudio inglés, éste retrospectivo, recuperó 10.267 mamografías que habían sido leídas por dos lectores y las sometió a una nueva lectura por otro radiólogo + CAD. Encontraron que la lectura única con CAD producía un 6,5 % más de cánceres con un aumento de la recitación que pasó de 6,5 % a 8,6 % ¹³¹. Otros investigadores compararon el efecto de la doble lectura respecto a la lectura simple con CAD en 6.381 mamografías de cribado y encontraron que la doble lectura detectó dos cánceres (11+2) que no habían sido detectados por la lectura simple con CAD. Es interesante remarcar que el CAD sí que marcó estos dos cánceres pero el radiólogo no consideró que las marcas correspondieran a ninguna lesión ¹³².

El estudio más relevante de este grupo es un estudio de equivalencia realizado por Gilbert *et al.* en el que 28.204 mamografías fueron leídas por tres radiólogos diferentes. En un brazo del estudio se siguió un protocolo de doble lectura y en el otro brazo un protocolo de lectura simple asistida con CAD. En el brazo de doble lectura se detectaron 199 de 227 cánceres (87,7 %) y en el brazo de lectura simple con CAD 198 de 227 (87.2 %) (P=0.89). La tasa de recitación fue de 3,4 % para la doble lectura y de 3,9 % para la lectura simple con

CAD. Esta diferencia es pequeña, pero significativa ($P < 0,001$). La sensibilidad, especificidad y valor predictivo positivo para la lectura única con CAD fueron de 87,2 %, 96,9 % y 18 %, respectivamente. Los valores para la doble lectura fueron de 87,7 %, 97,4 % y 21,1 %, respectivamente ¹³³.

ESTUDIOS LONGITUDINALES

Comparando lectura única con lectura única con CAD, Cupples *et al.* estudiaron el rendimiento de los radiólogos antes del CAD (7.872 mamografías) versus después del CAD (19.402 mamografías) y encontraron un aumento en la tasa de detección de cáncer del 16,1 % (de 3,7 % a 4,3 %) ¹³⁴, mientras que Gur *et al.* comparó 56.432 mamografías antes del CAD con 59.139 leídas con la ayuda del CAD en un contexto de mamografía asistencial en un hospital universitario y observó un aumento del 1,7 % ¹³⁵. Gromet revisó 112.413 mamografías que habían sido doblemente leídas y las comparó con 118.808 que fueron leídas por un radiólogo junto con el CAD. No observó diferencias significativas en la sensibilidad para la detección de cáncer ¹³⁶.

En la lectura de la mamografía de cribado, muchas áreas de potencial alteración son consideradas por el radiólogo como asimetrías glandulares o zonas de superposición de estructuras que crean una falsa imagen y, en consecuencia, no son consideradas tributarias de recitación. Por otra parte, en la práctica clínica diaria hay cierta presión para mantener la tasa de recitación lo más baja posible (idealmente por debajo del 5 %). Además, el número de marcas que pone el CAD es elevado y la frecuencia de cáncer muy baja. Todo ello contribuye a que en escenarios de práctica asistencial real el radiólogo pueda desestimar las marcas CAD.

A pesar de todo el esfuerzo de investigación dedicado al CAD, se puede concluir que la evidencia científica es insuficiente para determinar si la lectura simple con CAD dará resultados equivalentes a los de la doble lectura¹³⁷.

II-JUSTIFICACIÓN

El cribado poblacional de cáncer de mama es una intervención preventiva para disminuir el riesgo de morir de cáncer de mama. Los estudios aleatorizados poblacionales que se realizaron en las décadas de los 70 y 80, las revisiones y seguimiento de los mismos publicados más tarde y otros estudios observacionales han demostrado que la mamografía de cribado disminuye el riesgo relativo de morir por cáncer de mama en un 20 % en las mujeres invitadas a participar en un programa de cribado poblacional. En las mujeres que efectivamente participan en el mismo esta disminución es más importante, llegando en algunos casos al 40-50 % ^{13-18,23,29,31,34}.

Aunque la mamografía es la mejor prueba disponible para realizar la detección precoz del cáncer de mama, no está exenta de problemas. Es una prueba cuya sensibilidad en programas de cribado estaría, en el mejor de los casos, en torno al 85-90 % ya que hay una serie de cánceres de muy difícil visualización en la mamografía, especialmente en mamas densas y además, se ha constatado la existencia de una serie de errores dependientes tanto de la técnica como del radiólogo ^{65,71,75}.

El radiólogo puede cometer errores de detección y de interpretación. Para limitar estos errores, la guía europea para el cribado de cáncer de mama recomienda la doble lectura de la mamografía de cribado, al menos hasta que el rendimiento de los radiólogos haya sido totalmente evaluado ¹³⁸. Por otra parte, la doble lectura es un procedimiento costoso y limitado por la escasez de radiólogos expertos disponibles, de forma que en la práctica habitual de la mayoría de programas, y debido a la falta de suficientes radiólogos con dedicación completa a la mama, se suele emparejar un radiólogo de mama con un radiólogo general, con lo que la aportación es mayormente unidireccional, es decir, en la mayoría de ocasiones el lector experto contribuye a aumentar la sensibilidad del inexperto.

El CAD es una tecnología emergente basada en un software de reconocimiento de patrones que aplicado sobre una imagen mamográfica identifica hallazgos sospechosos. El CAD dibuja diversas marcas sobre la mamografía que atraen la atención del radiólogo con el objeto de disminuir los errores de detección. Por tanto, el CAD podría constituir un nuevo sistema de apoyo al radiólogo para la detección de cáncer de mama.

Múltiples estudios retrospectivos o en situaciones “de laboratorio” comentados en la introducción demuestran el beneficio potencial del CAD aunque es necesario realizar grandes estudios poblacionales para valorar la influencia real del CAD en el proceso de toma de decisión del radiólogo. Por esta razón, varios investigadores recomiendan la realización de nuevos estudios prospectivos que valoren el CAD en la práctica clínica de un programa de cribado poblacional. Por otra parte, mucho de lo publicado en la literatura se basa en imágenes de mamografía convencional posteriormente escaneadas con la merma de calidad que ello supone, por lo que sería deseable realizar nuevos estudios prospectivos utilizando mamografía digital directa como fuente de origen de las imágenes^{114,122,137,139}.

III- HIPÓTESIS Y OBJETIVOS

3.1 Estudio 1

3.1.1 Hipótesis

El CAD es capaz de señalar la mayoría de cánceres que se diagnostican en un programa de cribado poblacional y especialmente aquellos de pequeño tamaño.

3.1.2 Objetivos

- 1- Determinar la sensibilidad y especificidad del CAD en neoplasias invasivas ≤ 10 mm diagnosticadas en el programa de cribado poblacional para la detección precoz de cáncer de mama en el área de influencia del Hospital Clínic.

- 2- Valorar la influencia de diversos factores radiológicos en la sensibilidad del CAD para la detección del cáncer de mama de pequeño tamaño.

3.2 Estudio 2

3.2.1 Hipótesis

El CAD puede substituir a uno de los lectores de la pareja que realiza la doble lectura en un programa poblacional de detección precoz de cáncer de mama.

3.2.2 Objetivos

- 1- Comparar los principales indicadores de rendimiento del programa de detección precoz en el área de influencia del Hospital Clínic, usando un nuevo protocolo de lectura en el que uno de los lectores sea substituido por el CAD respecto a los resultados obtenidos históricamente mediante el protocolo clásico de doble lectura con arbitraje.

- 2- Valorar la aplicabilidad del nuevo protocolo de lectura.

IV- PRESENTACIÓN DE LOS ARTÍCULOS PUBLICADOS

Role of Computer-Aided Detection in Very Small Screening Detected Invasive Breast Cancers

Xavier Bargalló · Martín Velasco · Gorane Santamaría ·
Montse del Amo · Pedro Arguis · Sonia Sánchez Gómez

Published online: 7 November 2012
© Society for Imaging Informatics in Medicine 2012

Abstract This study aims to assess computer-aided detection (CAD) performance with full-field digital mammography (FFDM) in very small (equal to or less than 1 cm) invasive breast cancers. Sixty-eight invasive breast cancers less than or equal to 1 cm were retrospectively studied. All cases were detected with FFDM in women aged 49–69 years from our breast cancer screening program. Radiological characteristics of lesions following BI-RADS descriptors were recorded and compared with CAD sensitivity. Age, size, BI-RADS classification, breast density type, histological type of the neoplasm, and role of the CAD were also assessed. Per-study specificity and mass false-positive rate were determined by using 100 normal consecutive studies. Thirty-seven (54.4 %) masses, 17 (25 %) calcifications, 6 (8.8 %) masses with calcifications, 7 (10.3 %) architectural distortions, and 1 asymmetry (1.5 %) were found. CAD showed an overall sensitivity of 86.7 %

(masses, 86.5 %; calcifications, 100 %; masses with calcifications, 100 %; and architectural distortion, 57.14 %), CAD failed to detect 9 out of 68 cases: 5 of 37 masses, 3 of 7 architectural distortions, and 1 of 1 asymmetry. Fifteen out of 37 masses were hyperdense, and all of them were detected by CAD. No association was seen among mass morphology or margins and detectability. Per-study specificity and CAD false-positive rate was 26 % and 1.76 false marks per study. In conclusion, CAD shows a high sensitivity and a low specificity. Lesion size, histology, and breast density do not influence sensitivity. Mammographic features, mass density, and thickness of the spicules in architectural distortions do influence.

Keywords Breast neoplasm · Cancer detection · Computer-assisted detection

This work was supported by a grant from Sociedad Española de Radiología Médica (Beca SERAM 03_XBC_INV_SERAM_11).

X. Bargalló (✉) · M. Velasco · G. Santamaría · M. del Amo ·
P. Arguis
Department of Radiology (CDIC), Hospital Clínic de Barcelona,
C/Villarroel, 170,
08036 Barcelona, Spain
e-mail: xbarga@clinic.ub.es

M. Velasco
e-mail: mvelasco@clinic.ub.es

G. Santamaría
e-mail: gsanta@clinic.ub.es

M. del Amo
e-mail: mdelamo@clinic.ub.es

P. Arguis
e-mail: parguis@clinic.ub.es

S. Sánchez Gómez
Department of Radiology, Hospital Marqués Valdecilla,
C/Herrera Oria sn,
Santander, Spain
e-mail: smsanchezgomez@gmail.com

Introduction

Screening mammography has a sensitivity of 85–90 % in women 49–69 years old, which means that 10–15 % of tumors are missed on mammography [1, 2]. The existence of missed cancers has been documented in various studies. Among cancers that are missed and detected at subsequent screening, 25–40 % can be defined in non-blinded retrospective reviews and considered worthy of work-up or biopsy. Missed cancers are related to several reasons: low disease prevalence, breasts density and complexity, subtle findings, and radiologist fatigue [3].

After the US Food and Drug Administration approved the first commercially available computer-aided detection (CAD) system in June 1998, CAD systems have been increasingly used. In fact, from 2004 to 2008, the use of CAD in screening mammography in the USA increased by 91 %. In 2008, CAD was used in 74 % of all screening mammographic studies in the USA [4]. Several studies in the literature have shown that CAD increases detection rates up to 20 % [5–11]. However, much of these studies were

performed by using digitized images which can be a source of variability.

Recently, the study Digital Mammographic Imaging Screening Trial has shown that full-field digital mammography (FFDM) is more accurate than conventional screen-film mammography in women under the age of 50 years, women with dense breasts in mammography, and premenopausal or perimenopausal women [12]. The reported sensitivity of CAD in FFDM varies from 78 to 96 % [10, 13–15] but none of these studies was focused on very small, screening detected, invasive breast cancers.

This study was aimed to evaluate the performance of CAD in very small (less than or equal to 1 cm) invasive breast cancers using FFDM. Mammographic features of the lesion (following BI-RADS descriptors), BI-RADS classification, breast density, size, and histology were assessed in order to determine their influence on CAD.

Materials and Methods

Patient Selection

Institutional review board approval was obtained. We have retrospectively studied 68 invasive breast cancers with a size equal to or lesser than 1 cm diagnosed in women aged 49 to 69 from our population-based breast cancer screening program. Our population-based breast cancer screening program involves three districts in the city of Barcelona (Sants-Montjuic, Eixample esquerra, and Les Corts). Between June 2007 and May 2010, 40,139 women were screened and 235 breast neoplasms were diagnosed (detection rate of 5.85/thousand). Fifty-seven out of 235 (24.2 %) were ductal carcinoma in situ which were excluded because we had planned to assess the sensitivity of CAD in invasive breast cancer. The remaining 178 were invasive cancers. Sixty-eight out of 178 (38.2 %) were invasive cancers that fulfilled the inclusion criteria.

CAD (SecondLook, version 7.2, iCAD, Inc) with FFDM (Senographe 2000 D, General Electric Healthcare) were used. Only screening mammograms were included. Each case had a craniocaudal (CC) and mediolateral oblique (MLO) views of the breast with cancer at the time of cancer diagnosis.

Mammographic and Histological Variables

Mammographic pattern of all visible cancers was recorded. Mammographic lesion types were classified as follows: masses, calcifications, masses with calcifications, architectural distortions, and asymmetries. To describe and classify radiological findings, we adhered to the BI-RADS lexicon. The variables were retrospectively assessed by XB and MV, both subspecialized in breast imaging with more than 10 and

25 years of experience, respectively. When no agreement, a consensus was reached.

Masses were classified according to their morphology (round, oval, lobular, and irregular), margins (well-defined, microlobular, ill defined, obscured, and spiculated), and density (hyper-, iso-, and hypodensity compared with surrounding fibroglandular tissue).

Classification of calcifications was performed according to their morphology, the number of flecks per cluster and the size of the cluster. Regarding the morphology—punctate, amorphous, pleomorphic, linear, branching, or casting. When at least one fine linear or branching calcification was clearly present, the cluster was classified in the last subgroup. Regarding the number of flecks per cluster, less than 10, 10–30, and more than 30 were classified. Regarding the size of the cluster, less than 1 cm, 10–30 mm, and more than 30 mm were classified.

An architectural distortion was defined as the presence of distortion without an associated mass. The total size of the lesion in mammograms and the thickness of the spicules were measured. We classified the thickness of the spicules into thin (<1 mm), medium (1–2 mm), or thick (>2 mm) in order to compare with the role of the CAD.

A focal asymmetry was considered when an asymmetry could not be accurately identified as a true mass (lack of any border).

Mammographic breast density was determined following ACR criteria. Type 1, predominantly fatty breast (<25 % of dense tissue); type 2, scattered fibroglandular densities (25–50 % of dense tissue); type 3, heterogeneously dense breast (50–75 % of dense tissue), and type 4, extremely dense breast (>75 % of dense tissue).

Each lesion was retrospectively assigned a final BI-RADS category as follows: 4a, low suspicion for malignancy; 4b, intermediate suspicion of malignancy; 4c, moderate concern, but not classic for malignancy; and 5, highly suggestive of malignancy.

All patients underwent surgery. Pathological diagnosis was available for all lesions. Size and histological type of the neoplasm were collected. The invasive cancers were histopathologically categorized as follows: invasive ductal carcinoma, lobular carcinoma, and a miscellaneous group including colloid and tubular types.

Following recommendations from European guidelines, we performed double reading with arbitration of discordant readings of each screening mammogram. We have compared cases with a false-negative reading of one of the radiologists with role of the CAD.

CAD Mark Evaluation

The method we used was as follows: we retrospectively sent the raw data of the images to CAD system. The CAD system

processed the images and sent them to the PACS. We retrieved the CAD processed images from PACS and checked the marks. CAD operating point was set at “H,” which means maximum sensitivity with poor specificity. CAD used two types of marks: ovals for “masses” (including true masses, asymmetries, and architectural distortions) and rectangles for calcifications. Each CAD mark was classified as either a true-positive or a false-negative mark. True-positive marks indicated correctly a malignant lesion. False-positive marks were all other CAD marks. For cancers appearing as masses, true positive was considered if the center of the lesion fell within the mark. The same marking principle was used for architectural distortions and asymmetries. For calcifications, true positive was considered when the CAD mark involved the majority of the elements of the cluster.

CAD marking according to view was also recorded.

Statistical Analysis

The sensitivity of the CAD system was calculated as the number of cancers correctly marked divided by the total number of cancers. The specificity of the CAD system was calculated as the number of negative studies (mammograms without CAD marks) in 100 consecutive normal four-view studies from our screening program.

The influence of several variables such as mammographic features (following BI-RADS descriptors), BI-RADS classification, breast density, and histology, on CAD performance was evaluated using the Chi-square test. The influence of size on CAD performance was evaluated using “Mann–Whitney *U* test.”

Results

Sixty-eight cancers were found in women with a median age of 62. We found 37 (54.4 %) masses, 17 (25 %) calcifications, 6 (8.8 %) masses with calcifications, 7 (10.3 %) architectural distortions, and 1 asymmetry (1.5 %). The average tumor diameter of invasive cancers found on the basis of masses was 7.35 mm; calcifications, 5.24 mm; masses with calcifications, 7 mm, and architectural distortions, 6.6 mm.

CAD showed an overall sensitivity of 86.7 %. Sensitivities according to the mammographic appearance are shown in Table 1.

CAD detected 59 out of 68 cases. Thirty-five out of 59 were marked in both views, 13 out of 59 were marked in craniocaudal view, and 11 out of 59 were marked in mediolateral oblique view. CAD failed to detect 9 out of 68 cases whose characteristics are shown in Table 2. Five out of 37 were masses, 3 out of 7 were architectural distortions, and 1 out of 1 was an asymmetry.

Table 1 Sensitivity according to mammographic appearance of cancer

| Mammographic appearance | Sensitivity (% (detected/total)) |
|----------------------------|----------------------------------|
| Calcifications | 100 (17/17) |
| Masses with calcifications | 100 (6/6) |
| Masses | 86.5 (32/37) |
| Architectural distortions | 57.14 (4/7) |

Five masses were not detected by CAD. No association was seen among mass morphology or margins and ability of the CAD to mark. Fifteen out of 37 masses were hyperdense, and all of them were detected by CAD ($p=0.047$). One out of 37 was hypodense and was not detected by CAD. On the other hand, 16 out of 37 masses were marked in both views, 8 were marked in CC view (only one showed spicules), and 8 were marked in MLOBL view.

Calcifications were all detected by CAD. The groups of calcifications measured between few millimeters and several centimeters on the mammographic image, but the actual size of the infiltrating neoplasm in the surgical specimen measured always less than 1 cm. Fourteen out of 17 groups of calcifications showed a cluster larger than 1 cm and all of them contained more than ten flecks. Fourteen of 17 were marked in both projections, 2 out of 17 were marked in craniocaudal view, and 1 out of 17 was marked in mediolateral oblique view. Morphology of the flecks was amorphous in three cases, pleomorphic in nine, and linear or branching in five.

Regarding architectural distortions, CAD failed in three cases in which spicules were thin. Conversely, four cases in which CAD was able to detect, spicules were medium or thick.

Regarding the relationship between CAD and BI-RADS category, CAD failed predominantly in lesions that were labeled as BI-RADS 4a. Likewise, CAD did not fail any case classified as BI-RADS 4c and 5 (Table 3) although this association was not statistically significant.

There were no statistically significant differences when comparing size and breast density with role of the CAD (Table 4). The average size of the lesions detected by CAD was 6.7 mm (SD, 2.4 mm), whereas that of lesions missed by CAD was 6.4 mm (SD, 2.4 mm).

Analysis of histopathological diagnosis—invasive ductal carcinoma was the predominant type accounting for 58 cases. There were four infiltrating lobular carcinoma, five infiltrating tubular carcinoma, and one colloid carcinoma. CAD detected all of these ten carcinomas.

Specificity of the CAD was 26 %. The CAD false-positive rate was 1.76 marks per mammography.

We had 14 false-negative readings in this group of 68 small invasive neoplasms. Six cases were masses, four were calcifications, three were masses with calcifications, and

Table 2 Characteristics of cancers that were missed by CAD

| Mammographic appearance | Breast density | BI-RADS category | Size (mm) | Histology | Morphology | Margins | Density |
|--------------------------|----------------|------------------|-----------|-----------|---------------|----------------|---------|
| Mass | Dense | 4a | 6 | IDC | Lobular | Ill-defined | Iso |
| Mass | Fatty | 4b | 9 | IDC | Irregular | Spiculated | Low |
| Mass | Fatty | 4a | 9 | IDC | Irregular | Microlobulated | Iso |
| Mass | Fatty | 4a | 6 | IDC | Irregular | Ill-defined | Iso |
| Mass | Fatty | 4a | 8 | IDC | Irregular | Spiculated | Iso |
| Architectural distortion | Dense | 4b | 1 | IDC | Thin spicules | | |
| Architectural distortion | Dense | 4b | 5 | IDC | Thin spicules | | |
| Architectural distortion | Fatty | 4b | 7 | IDC | Thin spicules | | |
| Asymmetry | Fatty | 4a | 7 | IDC | | | |

Fatty ACR breast density types 1–2, *Dense* ACR breast density types 3–4

one was an architectural distortion. CAD marked 11 out of 14 of these neoplasms failing to mark two masses and the architectural distortion.

Discussion

Our results using FFDM in very small invasive breast cancers have shown an overall sensitivity of 86.7 % with a specificity of 26 % and 1.76 false-positive marks per case. Our sensitivity for detecting masses has been 86.5 %, which is slightly inferior to that reported in other series with 91–96 % of overall sensitivity [10, 15–17]. These differences are mainly related to the size of the lesions. In our series, the mean size of cancers correctly prompted by CAD was 6.7 mm while in the study by Kim et al. the mean size was 22 mm (sensitivity, 96 %) [10] and in the study by Sadaf et al. it was 12.4 (sensitivity, 91 %). For tumors measuring between 1 and 10 mm, Sadaf et al. reported a sensitivity of 84 % whereas The et al. reported a sensitivity of 89 % for these small tumors. Nevertheless, note should be made that ductal carcinomas in situ had been included in these studies [16, 17]. Other factors that could explain

these small differences of sensitivity are the image parameters in each mammography unit, the brand of the CAD system, including software version and operating point, etc. In our study, we decided to set the operating point in “high” because the study was aimed to assess the maximum CAD sensitivity despite this approach may be detrimental to specificity.

CAD has shown to have a greater sensitivity for calcifications than for masses. Our study has shown a 100 % sensitivity of CAD for detecting calcifications and masses with calcifications. This is in accordance to other studies on CAD in FFDM [10, 15, 17]. In view of these results, it could be thought that radiologists can avoid exhaustive search for calcifications, looking instead at areas with calcifications detected by CAD. However, in spite of our high performance, we recommend to be prudent because some studies have shown poor results of CAD in detecting amorphous calcifications [10, 17, 18]. Since our study was focused on invasive carcinoma, most of calcifications were pleomorphic, fine linear or branching calcifications and most clusters had more than 1 cm and contained more than ten elements, which accounts for our highest sensitivity. Four cases in our series had amorphous calcifications, and all of them were detected by CAD.

Table 3 Relationships among BI-RADS classification, role of CAD, and mammographic appearance

| | CAD sensitivity (%) | CAD mistakes | Masses | Calcifications | Masses with calcifications | Architectural distortions | Asymmetries |
|-------------------|---------------------|----------------|--------|----------------|----------------------------|---------------------------|-------------|
| BI-RADS 4a (n=26) | 80.8 | 5 ^a | 22 | 2 | 1 | 0 | 1 |
| BI-RADS 4b (n=23) | 83.3 | 4 ^b | 7 | 7 | 3 | 6 | 0 |
| BI-RADS 4c (n=7) | 100 | 0 | 3 | 2 | 1 | 1 | 0 |
| BI-RADS 5 (n=12) | 100 | 0 | 5 | 6 | 1 | 0 | 0 |

^a Two were related to masses only seen on one view

^b Three corresponding to architectural distortions

Table 4 Breast density and CAD

| | CAD sensitivity (%) | CAD failed | CAD-marked both views | CAD-marked CC view | CAD-marked MLOBL view |
|---------------------------------|---------------------|------------|-----------------------|--------------------|-----------------------|
| Breast density ACR 1 ($n=17$) | 88.2 | 2 | 11 | 3 | 1 |
| Breast density ACR 2 ($n=21$) | 81 | 4 | 9 | 5 | 3 |
| Breast density ACR 3 ($n=25$) | 88 | 3 | 13 | 5 | 4 |
| Breast density ACR 4 ($n=5$) | 100 | 0 | 2 | 0 | 3 |

Sensitivity ACR 1–2, 84.2 %. Sensitivity ACR 3–4, 90 %

CAD marked both views in 59 % of detected cases which is in contrast with the study by Vega et al. who reported that 74 % of lesions were marked on both views [19]. In our study, the degree of concordance between views is lesser due to the smaller size of invasive neoplasms in our group of patients, which make them more difficult to be detected by CAD.

Our results have demonstrated that mass density is an important factor influencing the ability of the CAD to detect breast cancer. All of our hyperdense masses were correctly prompted, and one hypodense mass was not marked. These results are in agreement with those of Ellis et al. who reported a difference of 11–25 % (depending on different brands of CAD system) between isodense and hyperdense lesions [20]. Conversely, morphology and margins of the masses do not show any relationship with CAD performance in this group of small neoplasms.

Architectural distortions are the most difficult mammographic pattern to be detected by the radiologist and also by CAD. We have shown a low sensitivity for these lesions (57 %) which is in agreement with the study by Baker et al. who reported a sensitivity of 49 % using older CAD versions [21]. We have analyzed the mammographic and pathological size and the thickness of the spicules to search for associations which could aid in determining what variables are important to detect these elusive lesions. We have not found any relation to the size of the lesion (either mammographic or pathological). However, it is worth noting that three cases with thin spicules were not detected by CAD while the rest of cases showing medium or thick spicules were detected. Although the number of patients in this subgroup is too small to draw conclusions, we think that further research could help to confirm this point.

Our data emphasize the fact that lesions categorized as BI-RADS 4c or five were all detected by CAD. Interestingly, lesions categorized as BI-RADS 4a had the lowest CAD sensitivity which is likely related to the fact that half of CAD mistakes in this subgroup happened in masses only seen in one view (one of them could easily be confused with a lymph node) which likely made the radiologist gave them BI-RADS 4a category.

Sensitivity for cancers presenting as masses for CAD applied to analogical mammography has been lower in dense breast compared with non-dense breast. However, recent studies in FFDM have shown that CAD sensitivity for detecting breast cancer is similar in both, dense and non-dense breast. Thus, in the study by Sadaf et al. no differences were found [17]. The study by Yang et al. found a higher, but nonsignificant, difference for detection of mass lesions in fatty breasts [15]. Some other studies have shown that detection rate for malignant lesions tends to decrease while increasing background density [22–24]. Our study has shown that background density does not affect detection. However, most of our patients had non-dense breast because they came from our population-based screening program, which includes mainly postmenopausal women. Moreover, small lesions are easier to depict in non-dense breasts which accounts for the greater proportion of non-dense breast in our patients group. Therefore, no definitive conclusions can be drawn on this subject.

Similar to previous studies, histopathological type of cancer did not have significant influence on CAD performance in our study. Our study shows a high proportion of tubular carcinoma which is thought to be related to the small size of cancers in our study, and all of them were marked by CAD.

The specificity and the rate of false-negative marks is similar to previous studies [19, 20]. The specificity of the CAD is low, which represents a potential drawback for using this technology. On the other hand, sensitivity is quite good and CAD has also been able to correctly mark 11 out of 14 cases that were read as negative by one of our readers. Although this is a promising data, it should be kept in mind that in a hypothetical scenario in which one single reader were assisted by CAD, the radiologist could be distracted by the number of false-positive marks and perhaps wouldn't recall a true positive mark. Moreover, Taylor et al., reported that 40 % of false-negative readings are due to decision errors (lesions detected but reader decides not to recall the patient for further study) [25].

A limitation of our study is the small sample size. The strengths of our study include assessment of CAD sensitivity based on mammographic features with adherence to BI-

RADS descriptors and terminology and the fact that we have only included invasive carcinomas from our population-based screening program.

In view of our results, CAD works well for detecting high-density masses and calcifications but fails to detect some isodense masses and very thin spicules. Moreover, the number of false-negative marks can be distracting for the radiologist. We think that future developments of this technology should attempt to correct these limitations.

In conclusion, CAD shows a high sensitivity and a low specificity in detecting very small invasive breast cancers from our population-based screening program. Lesion size (between 1 and 10 mm), histology, and breast density do not influence sensitivity. Mammographic features turn out to be an important factor affecting CAD sensitivity. Thus, all invasive cancers showing calcifications and all hyperdense masses were correctly prompted. Lesions assigned as BI-RADS 4c-5 and thick spicules in architectural distortions seem to be easier to detect by CAD.

Conflict of Interest None

References

- Warren Burhenne LJ, Wood SA, D'Orsi CJ, Feig SA, Kopans DB, O'Shaughnessy KF, Sickles EA, Tabar L, Vyborny CJ, Castellino RA: Potential contribution of computer-aided detection to the sensitivity of screening mammography. *Radiology* 215:554–562, 2000
- Mushlin AI, Kouides RW, Shapiro DE: Estimating the accuracy of screening mammography: a meta-analysis. *Am J Prev Med* 14:143–153, 1998
- Vyborny CJ, Giger ML, Nishikawa RM: Computer-aided detection and diagnosis of breast cancer. *Radiol Clin North Am* 38:725–740, 2000
- Rao VM, Levin DC, Parker L, Cavanaugh B, Frangos AJ, Sunshine JH: How widely is computer-aided detection used in screening and diagnostic mammography? *JACR J American College of Radiology* 7:802–805, 2010
- Birdwell RL, Bhandodkar P, Ikeda DM: Computer-aided detection with screening mammography in a university hospital setting. *Radiology* 236:451–457, 2005
- Morton MJ, Whaley DH, Brandt KR, Amrami KK: Screening mammograms: interpretation with computer-aided detection-prospective evaluation. *Radiology* 239:375–383, 2006
- Gromet M: Comparison of computer-aided detection to double reading of screening mammograms: review of 231,221 mammograms. *AJR Am J Roentgenol* 190:854–859, 2008
- Ko JM, Nicholas MJ, Mendel JB, Slanetz PJ: Prospective assessment of computer-aided detection in interpretation of screening mammography. *AJR Am J Roentgenol* 187:1483–1491, 2006
- Khoo LA, Taylor P, Given-Wilson RM: Computer-aided detection in the United Kingdom national breast screening programme: prospective study. *Radiology* 237:444–449, 2005
- Kim SJ, Moon WK, Kim SY, Chang JM, Kim SM, Cho N: Comparison of two software versions of a commercially available computer-aided detection (CAD) system for detecting breast cancer. *Acta Radiol* 51:482–490, 2010
- Freer TW, Ulissey MJ: Screening mammography with computer-aided detection: prospective study of 12,860 patients in a community breast center. *Radiology* 220:781–786, 2001
- Pisano ED, Gatsonis C, Hendrick E, Yaffe M, Baum JK, Acharyya S, Conant EF, Fajardo LL, Bassett L, D'Orsi C, Jong R, Rebner M: Diagnostic performance of digital versus film mammography for breast-cancer screening. *N Engl J Med* 353:1773–1783, 2005
- Baum F, Fischer U, Obenaus S, Grabbe E: Computer-aided detection in direct digital full-field mammography: initial results. *Eur Radiol* 12:3015–3017, 2002
- Skaane P, Kshirsagar A, Stapleton S, Young K, Castellino RA: Effect of computer-aided detection on independent double reading of paired screen-film and full-field digital screening mammograms. *AJR Am J Roentgenol* 188:377–384, 2007
- Yang SK, Moon WK, Cho N, Park JS, Cha JH, Kim SM, Kim SJ, Im JG: Screening mammography-detected cancers: sensitivity of a computer-aided detection system applied to full-field digital mammograms. *Radiology* 244:104–111, 2007
- The JS, Schilling KJ, Hoffmeister JW, Friedmann E, McGinnis R, Holcomb RG: Detection of breast cancer with full-field digital mammography and computer-aided detection. *AJR Am J Roentgenol* 192:337–340, 2009
- Sadaf A, Crystal P, Scaranelo A, Helbich T: Performance of computer-aided detection applied to full-field digital mammography in detection of breast cancers. *Eur J Radiol* 77:457–461, 2011
- Soo MS, Rosen EL, Xia JQ, Ghate S, Baker JA: Computer-aided detection of amorphous calcifications. *AJR Am J Roentgenol* 184:887–892, 2005
- Bolivar AV, Gomez SS, Merino P, Alonso-Bartolome P, Garcia EO, Cacho PM, Hoffmeister JW: Computer-aided detection system applied to full-field digital mammograms. *Acta Radiol* 51(10):1086–1092, 2010
- Ellis RL, Meade AA, Mathiason MA, Willison KM, Logan-Young W: Evaluation of computer-aided detection systems in the detection of small invasive breast carcinoma. *Radiology* 245:88–94, 2007
- Baker JA, Rosen EL, Lo JY, Gimenez EI, Walsh R, Soo MS: Computer-aided detection (CAD) in screening mammography: sensitivity of commercial CAD systems for detecting architectural distortion. *AJR Am J Roentgenol* 181:1083–1088, 2003
- Malich A, Fischer DR, Bottcher J: CAD for mammography: the technique, results, current role and further developments. *Eur Radiol* 16:1449–1460, 2006
- Brem RF, Hoffmeister JW, Rapelyea JA, Zisman G, Mohtashemi K, Jindal G, Disimio MP, Rogers SK: Impact of breast density on computer-aided detection for breast cancer. *AJR Am J Roentgenol* 184:439–444, 2005
- Ho WT, Lam PW: Clinical performance of computer-assisted detection (CAD) system in detecting carcinoma in breasts of different densities. *Clin Radiol* 58:133–136, 2003
- Taylor P, Given-Wilson R, Champness J, Potts HW, Johnston K: Assessing the impact of CAD on the sensitivity and specificity of film readers. *Clin Radiol* 59:1099–1105, 2004

4.1 Estudio 1

“Role of computer-aided detection in very small screening detected invasive breast cancers”

Xavier Bargalló, Martín Velasco, Gorane Santamaría, Montse del Amo, Pedro Arguis y Sonia Sánchez Gómez.

Journal of digital imaging. 2013; 26:572-577

IMPACT FACTOR: 1,2 QUARTIL 3

4.1.1 Resumen de los resultados del estudio 1

4.1.1.1 Descripción de la muestra

Se encontraron 68 tumores invasivos ≤ 1 cm en mujeres con una mediana de edad de 62 años. Estos cánceres mostraron los siguientes patrones mamográficos: 37 (54,4 %) nódulos, 17 (25 %) calcificaciones, 6 (8,8 %) nódulos con calcificaciones, 7 (10,3 %) distorsiones de la arquitectura y 1 (1,5 %) asimetría focal. El tamaño tumoral medio fue de 7,35 mm en nódulos; 5,24 mm en calcificaciones; 7 mm en nódulos con calcificaciones; y 6,6 mm en distorsiones de la arquitectura. En las Figuras 8 y 9 se puede observar la mamografía de algunas de estas pequeñas neoplasias junto con el comportamiento del CAD.

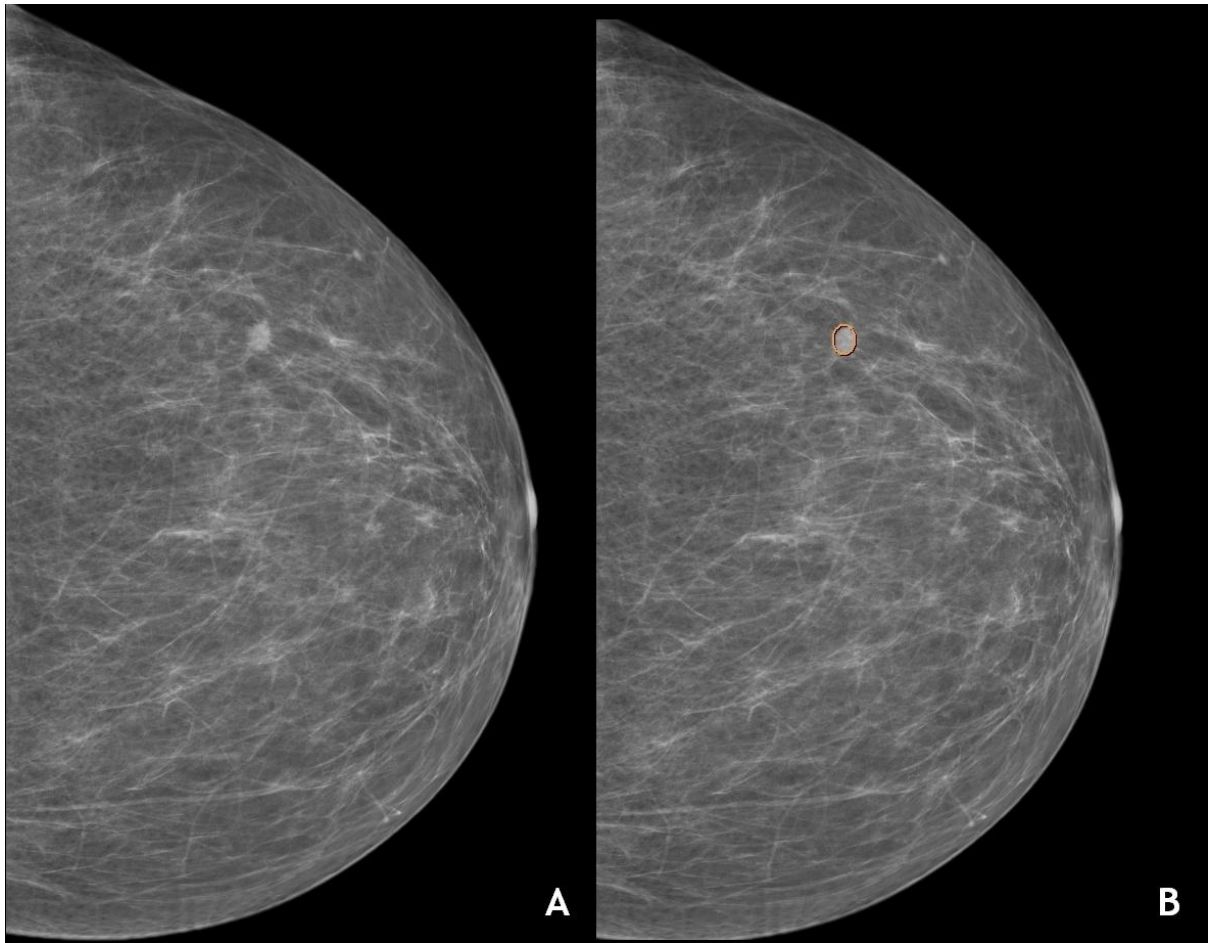


Figura 8. Proyección craneocaudal izquierda. A. Pequeña imagen nodular isodensa de márgenes irregulares. B. El CAD la marca correctamente. El estudio A-P demostró un carcinoma ductal infiltrante de 7 mm grado nuclear II.

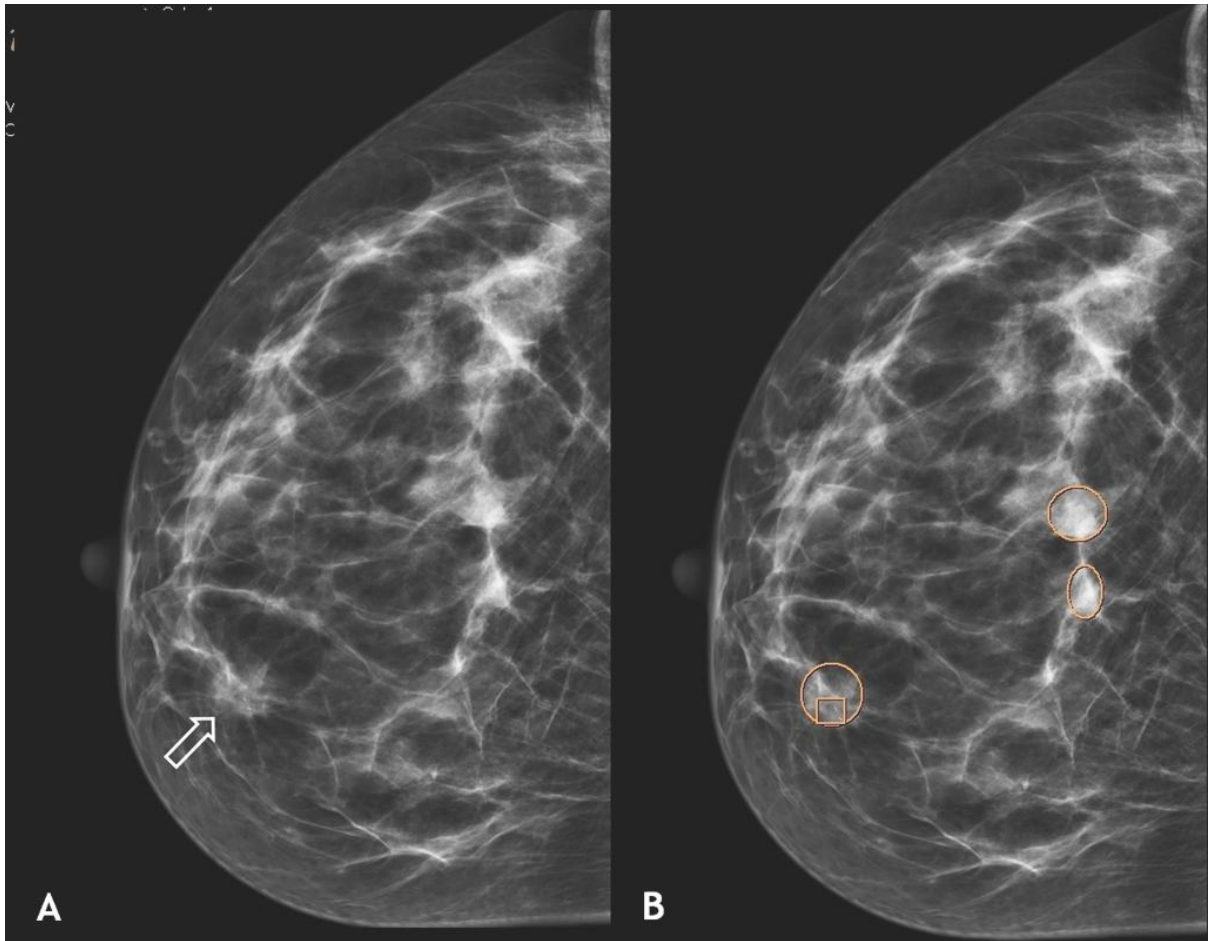


Figura 9. Proyección cráneo-caudal derecha. A. Pequeña asimetría focal iso-hipodensa (flecha) con algunas tenues calcificaciones e incipiente espiculación. B. El CAD marca correctamente en sus dos componentes (asimetría + calcificación) la lesión descrita. También se observan dos marcas falsas positivas en sectores posteriores de la mama. El estudio A-P demostró un carcinoma ductal *in situ* e infiltrante de 10 mm grado nuclear II.

4.1.1.2 Sensibilidad y especificidad del CAD

La sensibilidad del CAD para cáncer de mama infiltrante igual o menor a 10 mm fue de 86.7% y la especificidad fue del 26 % con una media de 1,76 marcas falsas positivas por mamografía.

El CAD detectó correctamente 59 de 68 tumores. Treinta y cinco de 59 fueron marcados en las dos proyecciones, 13 de 59 fueron marcados en la proyección craneocaudal y 11 de 59 fueron marcados en la proyección mediolateral oblicua. El CAD falló en la detección de 9 de 68 tumores cuyas características se muestran en la tabla 2. Cinco de 37 nódulos, 3 de 7 distorsiones de la arquitectura y la única asimetría focal que componía la muestra.

Tabla 2. Características de los cánceres que el CAD no marcó

| | Patrón radiológico | Densidad mamaria | BI-RADS | Tamaño (mm) | Histología | Morfología | Márgenes | Densidad |
|---|-------------------------|------------------|---------|-------------|------------|-----------------|----------------|-----------|
| 1 | Nódulo | Densa | 4a | 6 | CDI | Lobular | Mal definidos | Isodenso |
| 2 | Nódulo | Grasa | 4b | 9 | CDI | Irregular | Espiculados | Hipodenso |
| 3 | Nódulo | Grasa | 4a | 9 | CDI | Irregular | Microlobulados | Isodenso |
| 4 | Nódulo | Grasa | 4a | 6 | CDI | Irregular | Mal definidos | Isodenso |
| 5 | Nódulo | Grasa | 4a | 8 | CDI | Irregular | Espiculados | Isodenso |
| 6 | Distorsión arquitectura | Densa | 4b | 1 | CDI | Espículas finas | | |
| 7 | Distorsión arquitectura | Densa | 4b | 5 | CDI | Espículas finas | | |
| 8 | Distorsión arquitectura | Grasa | 4b | 7 | CDI | Espículas finas | | |
| 9 | Asimetría focal | Grasa | 4a | 7 | CDI | | | |

CDI: carcinoma ductal infiltrante

4.1.1.3 Factores que afectan el rendimiento del CAD

La tabla 3 muestra cómo afecta el patrón radiológico a la sensibilidad del CAD para detectar cáncer.

Tabla 3. Sensibilidad del CAD según el patrón radiológico

| Patrón radiológico | Sensibilidad (%) |
|-------------------------------|------------------|
| Calcificaciones | 100 |
| Nódulos con calcificaciones | 100 |
| Nódulos | 86.5 |
| Distorsión de la arquitectura | 57.1 |

Quince de 37 nódulos mostraron una densidad superior al parénquima glandular y todos ellos fueron detectados por el CAD ($p=0,047$). No se encontró asociación entre las variables morfología y contornos del nódulo y sensibilidad del CAD.

Los cánceres invasivos que se manifestaron como microcalcificaciones fueron todos detectados por el CAD. Los grupos de microcalcificaciones medían entre pocos milímetros y varios centímetros en la imagen mamográfica, pero el tamaño real del componente infiltrante en la pieza quirúrgica midió siempre menos de 1 cm. Catorce de los 17 grupos de microcalcificaciones mostraron un tamaño del grupo superior a 1 cm y todos ellos contenían más de 10 elementos. Catorce de 17 fueron correctamente marcados en las dos proyecciones, 2 de 17 solo se marcaron en la proyección craneocaudal y 1 de 17 en la mediolateral oblicua. La morfología de las microcalcificaciones fue amorfa en 3 casos, pleomórfica en nueve y linear o ramificada en cinco.

Las tres distorsiones que mostraban espícula fina no fueron detectadas por el CAD que en cambio sí detectó aquellas con espícula mediana o gruesa.

El tamaño medio de las lesiones detectadas por el CAD fue de $6,7 \pm 2,4$ mm y el de las lesiones no detectadas fue de $6,4 \pm 2,4$ mm lo que no constituye una diferencia significativa. La densidad de la mama y el tipo histológico del tumor no afectaron el rendimiento del CAD.

4.1.1.4 CAD y doble lectura

Catorce de las 68 neoplasias (20,6 %) que componen la muestra para este estudio habían sido un falso negativo de uno de los lectores de la pareja. Es decir, uno de los lectores o bien no vio la neoplasia o bien interpretó el hallazgo como benigno. Seis eran nódulos, cuatro grupos de microcalcificaciones, tres nódulos con microcalcificaciones y uno una distorsión de la arquitectura. El CAD marcó correctamente once de estas neoplasias fallando en dos de los nódulos y en la distorsión.



Single reading with computer-aided detection performed by selected radiologists in a breast cancer screening program



Xavier Bargalló^{a,*}, Gorane Santamaría^a, Montse del Amo^a, Pedro Arguis^a, José Ríos^c,
Jaume Grau^b, Marta Burrel^a, Enrique Cores^a, Martín Velasco^a

^a Department of Radiology (CDIC), Hospital Clínic de Barcelona, C/ Villarroel, 170, 08036 Barcelona, Spain

^b Preventive Medicine and Epidemiology Unit, Hospital Clínic de Barcelona, C/ Villarroel, 170, 08036 Barcelona, Spain

^c Biostatistics and Data Management Core Facility, IDIBAPS, (Hospital Clínic) C/ Mallorca, 183. Floor -1. Office #60. 08036 Barcelona, Spain

ARTICLE INFO

Article history:

Received 11 July 2014

Accepted 13 August 2014

Keywords:

Mammography

Breast screening

Breast neoplasms

Computer-assisted diagnosis

Breast

ABSTRACT

Objectives: To assess the impact of shifting from a standard double reading plus arbitration protocol to a single reading by experienced radiologists assisted by computer-aided detection (CAD) in a breast cancer screening program.

Methods: This was a prospective study approved by the ethics committee. Data from 21,321 consecutive screening mammograms in incident rounds (2010–2012) were read following a single reading plus CAD protocol and compared with data from 47,462 consecutive screening mammograms in incident rounds (2004–2010) that were interpreted following a double reading plus arbitration protocol. For the single reading, radiologists were selected on the basis of the appraisal of their previous performance.

Results: Period 2010–2012 vs. period 2004–2010: Cancer detection rate (CDR): 6.1% (95% confidence interval: 5.1–7.2) vs. 5.25%; Recall rate (RR): 7.02% (95% confidence interval: 6.7–7.4) vs. 7.24% (selected readers before arbitration) and vs. 3.94% (all readers after arbitration); Predictive positive value of recall: 8.69% vs. 13.32%. Average size of invasive cancers: 14.6 ± 9.5 mm vs. 14.3 ± 9.5 mm. Stage: 0 (22.3/26.1%); I (59.2/50.8%); II (19.2/17.1%); III (3.1/3.3%); IV (0/1.9%). Specialized breast radiologists performed better than general radiologists.

Conclusions: The cancer detection rate of the screening program improved using a single reading protocol by experienced radiologists assisted by CAD, at the cost of a moderate increase of the recall rate mainly related to the lack of arbitration.

© 2014 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Screening mammography reduces breast cancer mortality among invited women by 25% in incidence-based mortality studies and 31% in case-control studies (38% and 48%, respectively, among screened women) [1]. Although mammography is the most commonly used screening tool to detect breast cancer, its sensitivity is estimated at about 70–85% [2].

Double reading with arbitration is the standard procedure in most regional screening programs in Europe because it has been shown to increase the cancer detection rate (CDR) by 5–15% [3,4]. In general, an experienced breast radiologist is paired with a general diagnostic radiologist or even with a radiographer in some countries. The European guidelines for breast cancer screening programs recommend double reading, even though they also state that

“double reading is recommended in centralized programs for the first screening round and until the performance of the radiologists can be fully assessed” [5].

On the other hand, computer-aided detection (CAD) systems are becoming widely used [6]. Several studies in the literature show that CAD increases the detection rates by up to 20% [7,8]. The reported sensitivity of CAD in full-field digital mammography (FFDM) varies from 78 to 96% [9–11]. Gilbert et al. [12] demonstrated in a large study of equivalence that single reading with CAD could be an alternative to double reading.

Taking into account that European guidelines allow a single reading once the radiologist's performance has been fully assessed, we undertook this prospective study, in which mammograms of incident rounds were read by only one experienced radiologist assisted by CAD. The purpose of the study was to assess the impact of shifting from a standard double reading plus arbitration protocol to a new reading protocol consisting of a single reading with CAD done by radiologists selected for their good historical performance.

* Corresponding author. Tel.: +34 932275400; fax: +34 932275454.
E-mail address: xbarga@clinic.cat (X. Bargalló).

2. Materials and methods

2.1. Background

Our population-based breast cancer screening program started in May 2002 and involved the areas of Sants-Montjuic, Les Corts, and Eixample Esquerre in the city of Barcelona. The target population included women between 50- and 69-years-old who were invited to participate every 2 years. From 8 May 2002 to 16 November 2010, 85,304 mammograms were performed.

Following the recommendations of European guidelines [5], we performed a double reading plus arbitration of discordant readings by a third radiologist who was the most experienced of the team.

2.2. Study design

This was a prospectively designed study approved by the ethics committee. Informed consent was not required as the ethics committee asked for a surveillance protocol to be set up to ensure that the CDR did not drop to unacceptable levels and that acceptable recall rates (RR) were maintained.

From 17 November 2010 to 28 February 2013, 21,321 consecutive screening mammograms from women of the incident rounds were interpreted. All women from incident rounds had been screened at least once before in our program. The performance of the screening program was evaluated using a reading scheme based on a single reading plus CAD performed by radiologists selected according to their previous performance. There was a short period of 1 month to allow the readers to become familiar with CAD. Data from the prospective study were compared with those from 16 June 2004 (the date at which second round began) to 17 November 2010. In this period, 47,462 consecutive screening mammograms from incident rounds were performed. These mammograms followed a process of double reading with arbitration.

The following variables were collected: CDR; RR; biopsy rate; positive predictive value (PPV) of the recall; histological type; and stage of cancer detected.

2.3. Radiologist selection

Annual clinical audits were routinely performed. The performance of radiologists participating in the program was determined by collecting their historical results. The best radiologists were selected according to the CDR and RR.

The CDR in Catalonia in 2008–2009 was 3.91% (3.81% in incident rounds), with an RR of 4.5% [13]. In the city of Barcelona, the CDR in 2009 was 4.9% (4.4% in incident rounds) [14]. Based on these demographic data, we decided to set the inferior limit of CDR to 4.4% and the RR between 5 and 10%. Although European guidelines recommend that the RR be less than 5% in incident rounds, we decided to increase this limit to prioritize the detection as, in general, radiologists with a higher CDR tend to have a higher RR [15].

2.4. Mammograms and CAD system

All screening mammograms were performed using FFDM equipment (Senographe 2000 D, General Electric®). Standard two-view cranio-caudal and mediolateral oblique mammograms of each breast were done.

CAD (SecondLook, Premier, iCAD®) was used. The CAD operating point was set at "H", which means maximum sensitivity with poor specificity. CAD used two types of marks: ovals for masses (including true masses, asymmetries, and architectural distortions) and rectangles for calcifications.

The sensitivity of the CAD system was calculated as the number of cancers correctly marked divided by the total number of cancers. The specificity of the CAD system was calculated as the number of negative studies (mammograms without CAD marks) in 6674 consecutive mammograms. We also collected the American College of Radiology (ACR) type of the breast. All mammograms were read on a dedicated IDI MammoWorkstation (General Electric®).

2.5. Reading protocol

2.5.1. Prospective study (2010–2012)

Firstly, the selected radiologists read the mammograms in a standard way and made a decision as to whether to recall the patient. At this point, the CAD marks were engaged. The radiologists revised the marks and made a final decision according to the marks and their own opinion.

If the final decision was that the mammographic findings were BI-RADS (Breast Image Report and Data System) 1 or 2, it meant that CAD marks were not considered. Otherwise, if the finding was considered to be BI-RADS 0, 4, or 5, the patient was recalled and the radiologist wrote down the type of recall according to the following procedure:

- Type A recall: the radiologist found some mammographic finding that CAD did not mark
- Type B recall: the radiologist found some mammographic finding that CAD also marked
- Type C recall: the radiologist found the mammogram to be BI-RADS 1 or 2 but after considering the CAD marks, determined that a lesion could indeed be present and decided to recall. Therefore, CAD changed the radiologist's mind.

2.5.2. Historical cohort (2004–2010)

All screening mammograms were double read. The pairs were established by trying to match an experienced radiologist with a less experienced one. Both radiologists read blinded to the other's readings (independent double reading). When a major disagreement occurred between the readers (one reader recalled and the other did not), a third radiologist, the most experienced radiologist in the team, acted as a tiebreaker by assessing the case and making a final decision.

2.6. Monitoring of prospective study

As the CDR may vary slightly from 1 month to another, we decided to collect the CDR quarterly to obtain homogenous data. The 95% confidence interval (CI) was determined for each quarter and compared with the average historical value of that quarter. Feedback was provided to the readers. The results are shown in Fig. 1.

If the historical CDR for one quarter was above the upper CI limit of the same quarter in the prospective research, the study was stopped because it would mean that the radiologists were detecting fewer cancers compared with the historical period, which would have been unacceptable.

2.7. Statistical analysis

Evaluation of data was undertaken using Excel (Microsoft, 2010). The results are shown as absolute frequencies and rates by 100 or 1000 mammograms (CDR). For the estimations of 95% CI in the prospective study, the link between the Poisson and Chi-square distributions was used [16]. This method is based on the Poisson distribution and is more accurate and exact than methods based on the normal distribution. Uncommon events in populations, such as the occurrence of breast cancer are usefully modeled using the

Table 1
Overall performance of breast cancer screening program -on incident rounds-

| | Mammograms | Recalls | Recall rate (%) | Biopsies | Biopsy rate (%) | Cancers | C.D.R. (‰) | P.P.V. of recall (%) |
|--|------------|---------|-----------------|----------|-----------------|---------|------------|----------------------|
| 2010-2012: prospective study (single reading by selected radiologist plus CAD) | 21321 | 1496 | 7.02 | 218 | 1.02 | 130 | 6.10 | 8.69 |
| 2004-2010: historical period (double reading plus arbitration) | 47462 | 1870 | 3.94 | 429 | 0.90 | 249 | 5.25 | 13.32 |

C.A.D.: Computer-aided Diagnosis; P.P.V.: Positive Predictive Value

Table 2
Radiologists' performance 2004-2010 – historical period.

| Radiologist | Mammograms | Recalls | Recall rate (%) | Biopsies | Biopsy rate (%) | Cancers | C.D.R. (‰) | P.P.V. of recall (%) |
|-------------|------------|---------|-----------------|----------|-----------------|---------|------------|----------------------|
| A | 7903 | 524 | 6.63 | 55 | 0.70 | 28 | 3.54 | 5.34 |
| B | 6691 | 208 | 3.11 | 36 | 0.54 | 19 | 2.84 | 9.13 |
| C | 13779 | 1336 | 9.70 | 141 | 1.02 | 67 | 4.86 | 5.01 |
| D | 13038 | 726 | 5.57 | 102 | 0.78 | 64 | 4.91 | 8.82 |
| E | 12310 | 625 | 5.08 | 128 | 1.04 | 72 | 5.85 | 11.52 |
| F | 12246 | 554 | 4.52 | 81 | 0.66 | 39 | 3.18 | 7.04 |
| G | 9774 | 1063 | 10.88 | 76 | 0.78 | 38 | 3.89 | 3.57 |
| H | 10734 | 921 | 8.58 | 143 | 1.33 | 76 | 7.08 | 8.25 |
| I | 5126 | 186 | 3.63 | 22 | 0.43 | 14 | 2.73 | 7.53 |

Poisson distribution. The results of the historical period were considered as demographic data and compared with the data obtained in the prospective study.

3. Results

3.1. Overall performance of breast cancer screening program

During the 8-year study period, 68783 consecutive incident round screening mammograms were performed. **Table 1** summarizes data from the historical period and the prospective study. The CDR in the prospective study was 6.1‰ (95% CI: 5.1–7.2) vs. 5.2‰ in the historical period. The CDR increased by 16.2%. The RR in the prospective study was 7.02% (95% CI: 6.7–7.4) vs. 7.24 in the historical period for the selected readers before arbitration and vs. 3.94 for all readers after arbitration. Biopsy rate increased by 0.1%. P.P.V. of recall decreased by 4.63%.

3.2. Selection of radiologists and individual comparison between periods:

Only four out of our nine readers fulfilled the inclusion criteria (CDR>4.4‰ and RR5-10%) of the prospective study (**Table 2**). Radiologists E and H were breast radiologists with more than 5 years of experience. Radiologist C had worked previously in a breast unit

for more than 5 years. Radiologist D was the only one in the pool of screening radiologists who fulfilled the criteria without having previous experience as staff in a breast diagnostic unit.

Table 3 displays the individual comparison of selected radiologists between periods. Their individual RR decreased by 3.04%. In the prospective period was 7.02% (95% CI: 6.7–7.4) vs. 7.24 in the historical period. The three radiologists who were working (E, H) or had worked (C) in a breast unit managed to increase their CDR, reducing in some cases their personal RR. Radiologist D had a slight decrease in his CDR. The two specialized breast radiologists had a PPV of recall higher than 10%.

3.3. CAD performance:

CAD failed to detect 12 out of 230 cancers (sensitivity: 94.78%; specificity: 10.5%). Of these 12 cases, one was also a false negative of mammography because it was a sonographic finding. The remaining cases presented on mammography as a mass (2), calcifications (1, in a patient with breast implants), architectural distortion (5), asymmetry (2), and architectural distortion with subtle calcifications (1). Ten out of 12 cases were labeled by the radiologist as BI-RADS 0 and two as BI-RADS 4 (distortion with calcifications and calcifications in the patient with implants).

The recalls were classified as follows:

Type A recall 379, 2.9% (11/379) were cancer; Type B recall 1092, 10.8% (118/1092) were cancer; and Type C recall 25, none of them turned out to be cancer.

The average number of CAD marks per mammogram was 3.36 (2.13 masses and 1.23 calcifications).

3.4. Stage and histology of breast cancers

The breast cancer stages and histology of cancers detected in the two periods are shown in **Tables 4 and 5**, respectively. No statistical differences were found.

4. Discussion

The CDR of our screening program improved in the period 2010–2012 when the readings were made by selected radiologists plus CAD. CDR increased by 16.2% at the expense of increasing the

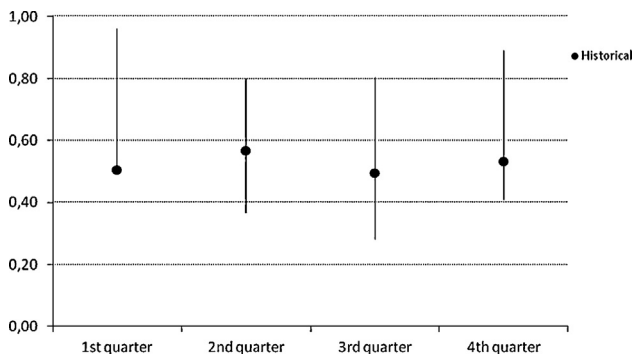


Fig. 1. Follow-up of prospective study. The y-axis indicates cancer detection rate in %. The x-axis indicates quarters of the year. Vertical lines indicate confidence interval (95%). The average historical value for each quarter is shown as a black point.

Table 3
Individual comparison of selected radiologists' performance between periods.

| Radiologist | RR: 4–10 (%) | RR: 10–12 (%) | BR: 4–10 | BR: 10–12 | CDR: 4–10 | CDR: 10–12 | PPVR: 4–10 | PPVR: 10–12 |
|-------------|--------------|---------------|----------|-----------|-----------|------------|------------|-------------|
| A | 9.7 | 8.37 | 1.02 | 1.16 | 4.86 | 6.15 | 5.01 | 7.35 |
| B | 5.57 | 7.19 | 0.78 | 0.93 | 4.91 | 4.73 | 8.82 | 6.58 |
| C | 5.08 | 5.73 | 1.04 | 0.92 | 5.85 | 6.27 | 11.52 | 10.94 |
| D | 8.58 | 6.48 | 1.33 | 1.05 | 7.08 | 7.18 | 8.25 | 11.08 |
| Overall | 7.24 | 7.02 | 1.03 | 1.02 | 5.60 | 6.10 | 7.73 | 8.69 |

Results of the historical period are given before arbitration.

R.R: Recall Rate; B.R: Biopsy Rate; C.D.R: Cancer Detection Rate; P.P.V.R: Positive Predictive Value of the recall.

Table 4
Comparison of breast cancer stages between periods.

| Stage | 0 | I | II | III | IV | Unknown |
|-----------|------------|-------------|------------|----------|----------|----------|
| 2010–2012 | 28 (21.5%) | 72 (55.4%) | 23 (17.7%) | 4 (3.1%) | 0 | 3 (2.3%) |
| 2004–2010 | 63 (25.3%) | 131 (52.6%) | 39 (15.8%) | 8 (3.2%) | 4 (1.6%) | 2 (0.8%) |

RR by 3 percentage points which was mainly due to the lack of arbitration. Radiologists specialized in breast imaging performed better than general radiologists. CAD sensitivity was 94.78% and specificity was 10.5%. CAD did not detect any cancer not perceived previously by the radiologist.

Several studies have evaluated the impact of CAD in a clinical setting using historical control methodology. Comparing single reading with single reading plus CAD, Cupples et al. [17] found that CDR increased by 16.1%, while Gur et al. [18] found an increase of only 1.7%. However, the 95% CI (–11 to 19%) reported in that study allowed for a wide range of detection rate changes. Similarly, Sánchez Gómez et al. [19] reported an increase of 1%, although radiologists dismissed some true positive marks. Gromet [20] reviewed 112,413 double read mammograms and compared them with 118,808 single read plus CAD mammograms. He found a small increase in sensitivity (but CDR decreased by 5.83%), with a lower RR. However, there was no arbitration in the double reading group. Gilbert et al. [12], in a large study of equivalence with matched-pair comparison, showed that CDR in the double reading group was 7.08%, while in the single reading plus CAD group it was 7.02% (a nonsignificant difference), opening the door to the introduction of CAD in population-based screening programs as a surrogate for one reader.

Our study is a historical control study in which single reading done by selected radiologists assisted by CAD was compared with double reading with arbitration. This approach was based on the following premises. First, many studies have supported the use of CAD in screening mammography [2]. Importantly, the study by Gilbert et al. [12] opened the door to start this kind of study. Second, we mindfully elected the radiologist team according to previous performance. The cut-off points were set at CDR >4.4‰ and RR 5–10%. Finally, a careful follow-up of indicators on a quarterly basis

Table 5
Comparison of breast cancer histology between periods.

| | 2010–2012 | 2004–2010 |
|--------------------------------|------------|-------------|
| Ductal carcinoma in situ | 28 (21.5%) | 63 (25.3%) |
| Invasive ductal carcinoma | 83 (63.8%) | 154 (61.8%) |
| Invasive lobular carcinoma | 11 (8.5%) | 17 (6.9%) |
| Mucinous carcinoma | 2 (1.5%) | 2 (0.8%) |
| Papillary carcinoma | 2 (1.5%) | 0 |
| Ductolobular carcinoma | 2 (1.5%) | 3 (1.2%) |
| Metaplastic carcinoma | 0 | 1 (0.4%) |
| Tubular carcinoma | 2 (1.5%) | 7 (2.8%) |
| Angiosarcoma of the breast | 0 | 1 (0.4%) |
| Primary lymphoma of the breast | 0 | 1 (0.4%) |

was performed to ensure that the CDR did not drop. Moreover, we collected the type of recall to be sure about the effect of CAD in the recalling process.

We had an increase by 16.2% in CDR. Although this increase did not reach statistical meaning, it can be considered relevant because it implies that up to 40 more cancers could have been detected in the historical period if the new reading system had been used. On the other hand, that increase in detection was not directly attributable to CAD because none of the readers recognized that CAD prompted a cancer not previously perceived by the radiologist. Thus, we recorded 25 Type C recalls and none of them turned out to be cancer. Therefore, we think that the improvement in the performance of the program might be mainly related to the improvement in the readers. However, comparing the selected readers before and after CAD introduction they had an improvement in their CDR of 8.9%, which could be related to several factors that are not easy to elucidate, i.e., the use of CAD could have influenced the reader positively, reinforcing the intention to recall in a subtle way; the arbitration, which has as a main function of reducing the recall rate, could have dismissed some images that, in the end, turned out to be cancer; and the reader could have paid more attention because he was reading alone (not double reading) or less attention because he trusted CAD.

Skaane et al. [21] stated that CAD may have the potential to increase cancer detection by up to 16%, and to reduce the number of interval cancers by up to 20%. However, they recognize that the true influence of CAD marks on decision-making could only be evaluated in a prospective clinical study. Our prospective study does not support such assumptions. Ergo, in the case that CAD had correctly prompted other cancers, our selected, high performing radiologists did not consider such marks as actionable. Therefore, in a real clinical background using breast radiologists, CAD did not provide cancers not detected by the radiologists (type C of recall did not result in any cancer). Further, Birdwell et al. [22], in a university hospital setting, concluded that CAD resulted in a 7.4% increase in cancers detected. Again, our results do not agree, which could be related to the fact that only two out of seven radiologists in that study were specialized in breast image. In return, three out of four of our radiologists were or had been breast radiologists, which usually have a higher CDR than general radiologists and also a higher reliability in their own judgment being less influenced by CAD marks.

Many studies have consistently demonstrated that the use of CAD is accompanied by an increase in the RR. Gilbert et al. [12] found a 15% increase in the recall rate and Cupples et al. [17] reported an 8.1% increase. Taylor and Potts [23] made a systematic

revision of the effect of CAD in screening mammography and concluded that the use of CAD increased the recall rate without having a significant effect on CDR. On the contrary, Gromet [20] stated that the RR decreased using single reading with CAD, even though he did not use arbitration in the double reading cohort. We found an increase of 3 percentage points in the recall rate, which is higher than that reported in similar studies. However, taking in consideration only the group of selected radiologists and comparing their RR before arbitration, we realized that, in fact, they have reduced slightly their personal RR. Therefore, we think that in this group of high performing radiologist, the CAD is not responsible for the increase in the overall RR, but the arbitration was responsible for the strong reduction of RR in the historical period.

Our breast cancer screening program includes clinical audit as part of quality assurance procedures to improve performance of radiologists. Moreover, on a yearly basis, readers are provided with feedback of their CDR, RR, and cancers missed. Likewise, challenging cases are shown and discussed to allow continued medical education. All of our radiologists read an average of 2700 screening mammograms per year. Several screening programs established a minimum of between 2000 and 5000 mammograms per year [24,25], although none of them have used their clinical audits as a means to select readers. Although all of our readers fulfilled the minimum of 2000 mammograms per year and were provided with feedback and continued medical education, only those with full dedication to breast imaging or very motivated to improve their performance were able to be included in the study. This is consistent with the study by Elmore et al. [15], who found that radiologists with fellowship training in breast imaging had better sensitivity but higher false-positive rates.

In our study, the cut-off points for selection were set at 4.4% for CDR and between 5 and 10% for RR. The study by Carney et al. [26] is one of the very few studies addressing the issue of identifying interpretative performance criteria for radiologists reading screening mammography. They found that the median performance (50th percentile) was a CDR of 3.7% and an RR of 9.2%. Further, the performance markers we found agree with those described by Sickles et al. [27], who reported a CDR of 6% for breast specialists and 3.4% for generalists.

We think that the main strength of this study is that it was prospectively designed. This allowed us to determine the role of CAD in recalls. Another good point is that we performed annual clinic audits, which allowed us to select the readers. To the best of our knowledge, this is the first paper in which radiologists have been selected according to their previous performance.

One limitation of this study is that there was not a unified regional tumor registry to establish the sensitivity and specificity. Consequently, we used the CDR and the RR as estimations of these variables.

As human behavior may be modified when people are scrutinized, readers' behavior could also be different when their process of reporting mammograms was registered. Thus, although none of the readers admitted that CAD had modified their intention to recall in cancers, it is possible that CAD had in fact influenced the radiologists. To clarify this point, it would have been necessary to not allow the radiologists to check the CAD marks until they had issued a report. This would have clearly separated the opinion of the radiologist with and without CAD. Further research may be necessary to investigate this issue.

To sum up, in centralized screening programs with an annual clinic audit, single reading by selected radiologists assisted by CAD could safely be incorporated in clinical practice of screening mammography at the cost of a moderate increase of the recall rate.

References

- [1] Paci E. Summary of the evidence of breast cancer service screening outcomes in Europe and first estimate of the benefit and harm balance sheet. *J Med Screen* 2012;19(1):5–13.
- [2] Birdwell RL. The preponderance of evidence supports computer-aided detection for screening mammography. *Radiology* 2009;253(1):9–16.
- [3] Thurjell EL, Lernevall KA, Taube AA. Benefit of independent double reading in a population-based mammography screening program. *Radiology* 1994;191:241–4.
- [4] Ciatto S, Ambrogetti D, Bonardi R, Catarzi S, Rizzo G, Rosselli Del Turco M, et al. Second reading of screening mammograms increases cancer detection and recall rates. Results in the Florence screening programme. *J Med Screen* 2005;12:103–6.
- [5] Perry N, Broeders M, de Wolf C, Törnberg S, Holland R, von Karsa L. European guidelines for quality assurance in mammography screening. 4th ed. Luxembourg: Office for Official Publications of the European Communities; 2006.
- [6] Rao VM, Levin DC, Parker L, Cavanaugh B, Frangos AJ, Sunshine JH. How widely is computer-aided detection used in screening and diagnostic mammography? *J Am Coll Radiol* 2010;7:802–5.
- [7] Freer TW, Ulissey MJ. Screening mammography with computer-aided detection: prospective study of 12,860 patients in a community breast center. *Radiology* 2001;220(3):781–6.
- [8] Baum F, Fischer U, Obenauer S, Grabbe E. Computer-aided detection in direct digital full-field mammography: initial results. *Eur Radiol* 2002;12(12):3015–7.
- [9] Yang SK, Moon WK, Cho N, Park JS, Cha JH, Kim SM, et al. Screening mammography-detected cancers: sensitivity of a computer-aided detection system applied to full-field digital mammograms. *Radiology* 2007;244(1):104–11.
- [10] Sadaf A, Crystal P, Scaranelo A, Helbich T. Performance of computer-aided detection applied to full-field digital mammography in detection of breast cancers. *Eur J Radiol* 2011;77(3):457–61.
- [11] Bargalló X, Velasco M, Santamaría G, Del Amo M, Arguis P, Sánchez Gómez S. Role of computer-aided detection in very small screening detected invasive breast cancers. *J Digit Imaging* 2013;26(3):572–7.
- [12] Gilbert FJ, Astley SM, Gillan MGC, Agbaje OF, Wallis MG, James J, et al. Single reading with computer-aided detection for screening mammography. *N Engl J Med* 2008;359:1675–84.
- [13] Asuncion N, Delfrade J, Salas D, Zubizarreta R, Ederra M. Breast cancer screening: characteristics and results of the Spanish programs. *Med Clin (Barc)* 2013;141(1):13–23.
- [14] Agència de Salut Pública de Barcelona. Butlletí Informatiu de l'Oficina Tècnica del Programa de Detecció Precoç de Càncer de Mama de Barcelona. Barcelona: Num.6: Agència de Salut Pública de Barcelona; 2011.
- [15] Elmore JG, Jackson SL, Abraham L, Miglioretti DL, Carney PA, Geller BM, et al. Variability in interpretive performance at screening mammography and associated with accuracy. *Radiology* 2009;253(3):641–51.
- [16] Armitage P, Berry G, Matthews JN. *Statistical methods in medical research*. Oxford: Blackwell Publishing; 2008.
- [17] Cupples TE, Cunningham JE, Reynolds JC. Impact of computer-aided detection in a regional screening mammography program. *Am J Roentgenol* 2005;185(4):944–50.
- [18] Gur D, Sumkin JH, Rockette HE, Ganott M, Hakim C, Hardesty L, et al. Changes in breast cancer detection and mammography recall rates after the introduction of a computer-aided detection system. *J Natl Cancer Inst* 2004;96(3):185–90.
- [19] Sanchez Gomez S, Torres Tabanera M, Vega Bolivar A, Sainz Miranda M, Baroja Mazo A, Ruiz Diaz M, et al. Impact of a CAD system in a screen-film mammography screening program: a prospective study. *Eur J Radiol* 2011;80(3):317–21.
- [20] Gromet M. Comparison of computer-aided detection to double reading of screening mammograms: review of 231,221 mammograms. *Am J Roentgenol* 2008;190(4):854–9.
- [21] Skaane P, Kshirsagar A, Hofvind S, Jahr G, Castellino RA. Mammography screening using independent double reading with consensus: is there a potential benefit for computer-aided detection. *Acta Radiol* 2012;53:241–8.
- [22] Birdwell RL, Bandokar P, et Ikeda DM. Computer-aided detection with screening mammography in a University Hospital Setting. *Radiology* 2005;236:451–7.
- [23] Taylor P, Potts HW. Computer aids and human second reading as interventions in screening mammography: two systematic reviews to compare effects on cancer detection and recall rate. *Eur J Cancer* 2008;44(6):798–807.
- [24] Cornford E, Reed J, Murphy A, Bennett R, Evans A. Optimal screening mammography reading volumes: evidence from real life in the East Midlands region of the NHS Breast Screening Programme. *Clin Radiol* 2011;66:103–7.
- [25] Rawashdeh MA, Lee WB, Bourne RM, Ryan EA, Pietrzyk MW, Reed WM, et al. Markers of good performance in mammography depend on number of annual readings. *Radiology* 2013;269(1):61–7.
- [26] Carney PA, Sickles EA, Monsees BS, Bassett LW, Brenner RJ, Feig SA, et al. Identifying minimally acceptable interpretive performance criteria for screening mammography. *Radiology* 2010;255(May (2)):354–61.
- [27] Sickles EA, Wolverton DE, Dee KE. Performance parameters for screening and diagnostic mammography: specialist and general radiologists. *Radiology* 2002;224:861–9.

4.2 Estudio 2

“Single reading with computer-aided detection performed by selected radiologists in a breast cancer screening program “

Xavier Bargalló, Gorane Santamaría, Montse del Amo, Pedro Arguis, José Ríos, Jaume Grau, Marta Burrel, Enrique Cores y Martín Velasco.

European journal of radiology. 2014; 83: 2019–2023

IMPACT FACTOR: 2,16 QUARTIL 2

4.2.1 Resumen de los resultados del estudio 2

4.2.1.1 Rendimiento global del programa de cribado

Durante el periodo de 8 años que duró el estudio se realizaron 68.783 mamografías de cribado consecutivas en mujeres en segunda ronda o sucesivas. La Tabla 4 muestra de forma comparativa los principales indicadores de rendimiento del programa en el periodo histórico (2004-10) respecto el periodo 2010-12 en que se realizó el estudio prospectivo.

Tabla 4. Rendimiento del programa de cribado -segundo cribado o sucesivos-

| | Mx | R | TR (%) | Biopsias | TB (%) | Cánceres | TDC (‰) | VPPR (%) |
|--|-------|------|--------|----------|--------|----------|---------|----------|
| 2004-10: Doble lectura con arbitraje | 47462 | 1870 | 3,94 | 429 | 0,90 | 249 | 5,25 | 13,32 |
| 2010-12: Lectura única con CAD | 21321 | 1496 | 7,02 | 218 | 1,02 | 130 | 6,10 | 8,69 |

Mx: mamografías; R: recitaciones; TR: tasa de recitación; TB: tasa de biopsia; TDC: tasa de detección de cáncer; VPPR: valor predictivo positivo de la recitación

La tasa de detección en el estudio prospectivo fue de 6,1 ‰ (IC 95%: 5,1-7,2) versus 5,25 ‰ en el periodo histórico. Por tanto, se observa un incremento del 16,2 %. La tasa de recitación en el estudio prospectivo fue de 7,02 % (IC 95%: 6,7-7,4) versus 7,24 % (lectores seleccionados antes del arbitraje) y versus 3,94 % (todos los lectores después del arbitraje) del periodo histórico. La tasa de biopsia aumentó en un 0,1 % y el valor predictivo positivo de la recitación disminuyó en un 4,63 %.

4.2.1.2 Rendimiento del CAD

El CAD falló en detectar 12 de los 230 cánceres que se diagnosticaron en el periodo 2010-2012 (sensibilidad del 94,78 % y especificidad del 10,5 %). La mayoría de fallos fueron distorsiones de la arquitectura (n=5), nódulos (n=2), calcificaciones (n=1) en una mujer portadora de implantes mamarios, asimetrías (n=2) y distorsión con sutiles calcificaciones (n=1).

El tipo de recitación que surge de la interacción entre el CAD y el radiólogo lector se muestra en la tabla 5 teniendo en cuenta que en el tipo A, el radiólogo ve una supuesta lesión y el CAD no marca nada. En el tipo B, el radiólogo ve una supuesta lesión y el CAD la marca. En el tipo C, el radiólogo no ve alteraciones pero el CAD marca una supuesta lesión. El radiólogo después de valorar la marca cambia su opinión inicial y recita a la paciente para estudio.

Tabla 5. Tipos de recitaciones surgidos de la interacción con el CAD

| Tipo de recitación | Número | Cánceres |
|--------------------|--------|--------------|
| Tipo A | 379 | 11 (2,9 %) |
| Tipo B | 1092 | 118 (10,8 %) |
| Tipo C | 25 | 0 |

Así, pues, el CAD no detectó ningún cáncer que no hubiese sido previamente visto por el radiólogo lector. El número medio de marcas falsas por mamografía fue de 3,36 (2,13 marcas tipo nódulos y 1,23 calcificaciones).

4.2.1.3 Estadificación e histología de los tumores detectados

Las tabla 6 muestra la estadificación de los tumores. No se observaron diferencias significativas entre periodos ni en cuanto la estadificación ni en cuanto a la histología de los tumores. El tamaño medio de los cánceres invasivos fue de $14,6 \pm 9,5$ mm en el periodo prospectivo versus $14,3 \pm 9,5$ mm en el histórico.

Tabla 6. Estadificación de los cánceres detectados

| | 0 | I | II | III | IV | Desconocido |
|------------------|------------|-------------|------------|----------|----------|-------------|
| 2004-2010 | 63 (25,3%) | 131 (52,6%) | 39 (15,8%) | 8 (3,2%) | 4 (1,6%) | 2 (0,8%) |
| 2010-2012 | 28 (21,5%) | 72 (55,4%) | 23 (17,7%) | 4 (3,1%) | 0 | 3 (2,3%) |

4.2.1.4 Rendimiento de los radiólogos

Como se observa en la tabla 7, de los nueve radiólogos que formaban el grupo lector en el periodo histórico, cuatro cumplieron los criterios de inclusión (TDC > 4,4 ‰ y TR: 5-10 %) para el nuevo protocolo de lectura (lector único + CAD). Los radiólogos E y H eran radiólogos de la unidad asistencial de mama con más de 5 años de experiencia y el radiólogo C había trabajado como adjunto en una unidad asistencial de mama.

Tabla 7. Rendimiento de los radiólogos en el periodo 2004-2010 –segundo cribado o sucesivos-

| Radiólogo | Mx | R | TR (%) | Biopsias | TB (%) | Cánceres | TDC (‰) | VPPR (%) |
|-----------|-------|------|--------|----------|--------|----------|---------|----------|
| A | 7903 | 524 | 6,63 | 55 | 0,70 | 28 | 3,54 | 5,34 |
| B | 6691 | 208 | 3,11 | 36 | 0,54 | 19 | 2,84 | 9,13 |
| C | 13779 | 1336 | 9,70 | 141 | 1,02 | 67 | 4,86 | 5,01 |
| D | 13038 | 726 | 5,57 | 102 | 0,78 | 64 | 4,91 | 8,82 |
| E | 12310 | 625 | 5,08 | 128 | 1,04 | 72 | 5,85 | 11,52 |
| F | 12246 | 554 | 4,52 | 81 | 0,66 | 39 | 3,18 | 7,04 |
| G | 9774 | 1063 | 10,88 | 76 | 0,78 | 38 | 3,89 | 3,57 |
| H | 10734 | 921 | 8,58 | 143 | 1,33 | 76 | 7,08 | 8,25 |
| I | 5126 | 186 | 3,63 | 22 | 0,43 | 14 | 2,73 | 7,53 |

Mx: mamografías; R: recitaciones; TR: tasa de recitación; TB: tasa de biopsia; TDC: tasa de detección de cáncer; VPPR: valor predictivo positivo de la recitación

V- DISCUSIÓN

5.1 Discusión de los resultados del estudio 1:

“Role of computer-aided detection in very small screening detected invasive breast cancers”

La sensibilidad del CAD para detectar cáncer invasivo ≤ 1 cm ha sido de 86,7 %, que está dentro del rango 84-89 % que son los valores publicados en la literatura para tumores de este tamaño ^{140,141,142}. No obstante, nuestra sensibilidad real es mayor ya que a diferencia de estos estudios, nuestra serie solo incluyó tumores invasivos. Dado que el carcinoma in situ se manifiesta en la mayoría de casos por calcificaciones, y el CAD las detecta muy bien, las series que incluyen carcinomas in situ tienden a mostrar sensibilidades más altas.

La sensibilidad del CAD para tumores de cualquier tamaño descrita en la literatura es ligeramente superior y muestra una relación lineal con el tamaño medio. Así, Kim *et al.* encontraron una sensibilidad del 96 % para tumores con un tamaño medio de 22 mm ¹¹⁶ y Sadaf *et al.* encontraron una sensibilidad del 91 % para un tamaño tumoral medio de 12,4 mm ¹⁴².

El CAD es mucho más sensible para detectar calcificaciones que el resto de manifestaciones radiológicas del cáncer de mama. En este estudio, todos los cánceres que se manifestaron como calcificaciones u otro patrón combinado con calcificaciones fueron detectados por el CAD, lo que concuerda con otros artículos que también han usado mamografía digital directa (DR) ^{116,140,142}. En vista de este resultado podríamos llegar a la errónea conclusión de que el radiólogo puede delegar la búsqueda de microcalcificaciones al CAD. Sin embargo, y a pesar de estos datos, la recomendación es prudencia ya que algunos estudios han mostrado pobres resultados del CAD para detectar calcificaciones amorfas ¹¹⁷. Además, conviene tener en cuenta que si hubiéramos calculado la sensibilidad por imagen, ésta no

llegaría al 100 % ya que si bien 14 de 17 casos con patrón cálcico fueron marcados por el CAD en ambas proyecciones, hubo tres que solo fueron marcados en una proyección. Este porcentaje sigue siendo mucho mejor que para las lesiones nodulares ya que 16 de 37 fueron marcadas en las dos proyecciones pero el resto no.

Weigel *et al.* publicaron que un 32 % de cánceres de mama se manifiestan únicamente como calcificaciones, incluyendo formas in situ e invasivas. En un trabajo previo de nuestro grupo encontramos que un 24 % de cánceres invasivos de menos de 1 cm se manifestaban como calcificaciones y un 6,5 % en forma de patrón mixto con nódulos y calcificaciones. La mayoría de patrones cálcicos estaban formados por grupos de más de 1 cm de extensión que contenían más de 10 microcalcificaciones, con una morfología predominantemente lineal o pleomórfica¹⁰⁶. Cuanto más pequeño es el tumor infiltrante, es más probable que se manifieste únicamente como un grupo de calcificaciones¹⁴³, lo que es lógico habida cuenta de la dificultad de detectar nódulos de menos de 5 mm excepto en mamas muy adiposas. Otros autores también han mostrado que la probabilidad de invasión aumenta en grupos con más de 10 calcificaciones y más de 10 mm de extensión. Así pues, el riesgo de invasión en grupos de menos de 10 elementos, 10-40 elementos y más de 40 elementos es de 15, 24 y 43 % respectivamente ($P < 0,05$)^{144,145}.

Las calcificaciones son la manifestación más frecuente del carcinoma in situ. La mayoría de nuestros patrones cálcicos, al ser la expresión mamográfica de carcinomas in situ con focos de infiltración, tenían un mayor número de elementos, lo que también explica nuestra elevada sensibilidad. Por otra parte, nuestra serie tiene 4 casos con calcificaciones amorfas, y todos ellos fueron correctamente detectados.

El CAD marcó correctamente la neoplasia en ambas proyecciones en 59 % de los casos, lo que contrasta con el estudio de Bolivar *et al.* que encontraron que un 74 % de sus neoplasias habían sido marcadas en ambas proyecciones¹⁴⁶. En nuestro estudio, el grado de concordancia entre proyecciones es menor, probablemente debido al menor tamaño de las neoplasias incluidas, lo que hace que sean más difíciles de detectar por el CAD.

Nuestro estudio demostró que la densidad de la lesión es un factor importante que influye en la capacidad del CAD para detectar las lesiones. Todas las lesiones nodulares que se catalogaron como hiperdensas fueron correctamente señaladas y la única que se definió como hipodensa no fue señalada, lo que concuerda con lo reportado en la literatura¹¹⁵. Otras características de la lesión como su morfología o márgenes no mostraron relación estadísticamente significativa con el rendimiento del CAD.

La distorsión de la arquitectura es el patrón más difícil de detectar tanto para el radiólogo como para el CAD. En este estudio, la sensibilidad del CAD para este grupo de lesiones fue de 57 %, que es baja y está en consonancia con lo publicado por Baker *et al.*, que mostraron una sensibilidad del 49 % usando versiones de CAD más antiguas⁷⁵. En nuestro estudio analizamos 7 distorsiones valorando factores como el tamaño mamográfico de la distorsión frente al anatomopatológico y el grosor de las espículas con el objeto de encontrar elementos distintivos que pudieran explicar cuáles de estas lesiones son más o menos accesibles para el CAD. No se encontró relación respecto al tamaño. No obstante, vale la pena mencionar que tres de los casos en los que la espícula era fina no fueron marcados y sí lo fueron el resto, en que la espícula era mediana o gruesa.

Los datos de este estudio enfatizan el hecho de que todas las lesiones categorizadas como BI-RADS 4-C o 5 fueron detectadas por el CAD. Es interesante notar que las lesiones BI-

RADS 4-A fueron las que más falló el CAD, lo que puede estar causado porque la mitad de lesiones en este grupo solo se veían en una proyección.

Es conocido que la mama densa tiende a ocultar lesiones que pueden resultar ser cáncer y, de hecho, la sensibilidad del CAD en mamas densas es menor. A partir del estudio de Pisano *et al.*, sabemos que en la mujer pre o perimenopáusica, o con mama densa la mamografía digital es superior a la analógica⁶⁴. Varios estudios han mostrado que la tasa de detección de lesiones malignas tiende a disminuir cuando aumenta la densidad mamaria^{147,148,149}. Nuestro estudio no muestra cambios en la detección relacionados con la densidad mamaria aunque hay que tener en cuenta que incluye pacientes en su mayoría postmenopáusicas y con predominio de mama grasa, lo que facilita la detección de pequeños tumores.

De forma similar a estudios previos, el tipo histológico de cáncer no influyó en el rendimiento del CAD en nuestro estudio. El carcinoma tubular se halló en una proporción mayor a lo esperable, lo que se relaciona con el pequeño tamaño de los tumores en nuestra serie, y todos ellos fueron marcados por el CAD.

Tanto la especificidad como las marcas falsas positivas fueron similares a estudios previos^{75,115}. La especificidad del CAD es baja, lo que representa un inconveniente para la implantación de esta tecnología en la práctica clínica. Por otra parte, la sensibilidad para la demostración de lesiones ≤ 1 cm es alta y dentro del rango de las publicadas en la literatura. Vale la pena incidir en el hecho de que 14/68 (21%) de estas neoplasias de pequeño tamaño habían sido falsos negativos de uno de los lectores, incluyendo varios casos que contenían calcificaciones muy tenues. El CAD fue capaz de marcar correctamente 11 de estos 14 casos (79%) lo que reafirma la potencialidad del CAD como instrumento de ayuda al radiólogo y eventualmente como sustituto de uno de los lectores de la pareja. Aunque es un dato

prometedor, hay que tener en cuenta que es un dato teórico, es decir que en la práctica diaria es posible que el lector, distraído o sobrepasado por la gran cantidad de marcas falsamente positivas (alrededor de un 80 % de las mamografías tienen marcas CAD que en realidad no corresponden a ninguna lesión), no prestara suficiente atención a marcas que eventualmente podrían ser un cáncer ¹⁵⁰. A esto hay que añadir que hasta un 40 % de las lecturas falsamente negativas se deben a errores de decisión (interpretación) que no están cubiertos por el CAD ¹⁵¹. Por este motivo, es crucial realizar estudios con CAD en la práctica clínica que incluyan un elevado volumen de pacientes del cribado poblacional.

Una limitación de este estudio es el pequeño tamaño de la muestra que es debido a la dificultad de diagnosticar tumores invasivos de menos de 1 cm de tamaño. No obstante, la proporción de tumores invasivos de menos de 1 cm en nuestro programa es del 36,3 % que está por encima del 30 % recomendado en la guía europea para el cribado de cáncer de mama ¹³⁸.

Las fortalezas de este estudio incluyen: a) la valoración de los hallazgos mamográficos siguiendo el léxico y criterios propuestos por el sistema BI-RADS, que es un estándar internacional en la descripción de las lesiones mamográficas, b) todos los tumores proceden del programa de cribado poblacional y c) todas las mamografías se realizaron con un mamógrafo digital directo.

Azavedo *et al.*, en una reciente revisión sistemática, afirman que la evidencia científica es insuficiente para determinar si la lectura simple con CAD es tan precisa como la doble lectura, indicando que son necesarios nuevos estudios en programas poblacionales que utilicen mamografía digital. Todos estos motivos nos condujeron a iniciar el estudio número dos de esta tesis ¹³⁷.

5.2 Discusión de los resultados del estudio 2:

“Single reading with computer-aided detection performed by selected radiologists in a breast cancer screening program”

Varios estudios han evaluado el impacto del CAD en la práctica clínica mediante estudios longitudinales también llamados de comparación histórica. En ellos se evalúa el comportamiento de unas variables en una población determinada antes y después de la introducción de una nueva técnica, medicación, instrumento, etc, con la hipótesis de trabajo de que la introducción de este factor tiene que mejorar el rendimiento pre-existente. En general, se considera que estos estudios son más rigurosos que los transversales aunque también tienen varios inconvenientes. Por una parte, existe un período de prueba o adaptación a la nueva técnica o sistema de trabajo que es difícil de cuantificar cuánto puede durar y que afecta a los resultados (por ejemplo la curva de aprendizaje). En segundo lugar y más importante, es que es muy difícil controlar si existen otras variables nuevas o diferentes que están afectando al desarrollo del estudio ¹²⁷.

Nuestra investigación es un estudio de control histórico en el que se ha introducido una intervención que es doble: por una parte, el CAD como asistente a la lectura radiológica y sustituto de un lector y, por otra, la selección de los radiólogos que pueden hacer la lectura única según su rendimiento histórico.

Este enfoque se basó en las siguientes premisas. Primero, existe abundante bibliografía dando validez al uso del CAD en la mamografía de cribado, aunque son principalmente estudios americanos cuyos programas de cribado poblacional son diferentes de los europeos. De forma especial, hay que señalar la investigación de Gilbert *et al.* que, en un gran estudio de equivalencia, demostraron que la tasa de detección de cáncer en el grupo

de doble lectura fue de 7,08 % mientras que en el brazo de lectura simple con CAD fue de 7,02 %¹³³, abriendo la puerta a la introducción del CAD en programas europeos de cribado poblacional de cáncer de mama como un sustituto del segundo lector. Segundo, una de las ventajas de los programas estructurados y centralizados de cribado poblacional es que tienen que adherirse a las guías europeas y, por tanto, se exige una exhaustiva auditoría anual de los resultados entre los que se mide el rendimiento de cada radiólogo lector. Esta auditoría nos permitió realizar una cuidadosa elección de los radiólogos que podían participar en nuestro estudio prospectivo de acuerdo a su rendimiento histórico. Finalmente, se realizó un estricto seguimiento de los indicadores de rendimiento del programa de forma trimestral para asegurar que la tasa de detección de cáncer no se viera perjudicada por el nuevo protocolo de lectura. Además, en cada recitación se registró el papel del CAD para asegurar cuál era su papel en el proceso diagnóstico.

Hemos encontrado un aumento del 16,2 % en la detección de cáncer que estaría en la línea de lo descrito por Cupples *et al.*¹³⁴ aunque a diferencia de éste, nuestro incremento no puede ser directamente atribuido al CAD ya que ninguno de los lectores reconoció que el CAD hubiese marcado un cáncer no visto por ellos previamente. Se recogieron 25 recitaciones tipo C que son aquellas en las que la paciente fue recitada por indicación del CAD; es decir, la marca del CAD hizo cambiar realmente la opinión del radiólogo que después de valorar la marca del CAD revaloró el caso y decidió recitar. Pero ninguna de estas 25 resultó ser un cáncer. Por todo ello, a nuestro juicio, la mejora en la detección de cáncer estuvo relacionada más bien con la selección de los radiólogos. Creemos que la aportación del CAD en nuestro estudio fue discreta, lo que estaría en concordancia con lo reportado en otros estudios longitudinales^{129,135}. No obstante, comparando los mismos lectores antes y después de la introducción del CAD se constató que había habido una

mejora de la tasa de detección de cáncer del 8,9 % que no es significativa. Este ligero aumento en el aumento de detección del radiólogo aunque de ningún modo es significativo, sí que podría estar en relación con varios factores difíciles de dilucidar, como pueden ser: a) el uso del CAD pudo haber influido de forma sutil al radiólogo, de forma que aunque éste no reconociera la ayuda, ésta efectivamente se produjese, b) los propios radiólogos al saber que no tenían la seguridad de una segunda lectura y que estaban siendo observados (dentro del estudio) pudieron haber estado más alerta, aunque este fenómeno es difícil que se mantenga durante los dos años que duró la fase prospectiva, c) el arbitraje, cuya misión principal es reducir la recitación, pudo haber descartado algunas imágenes que finalmente resultaron ser cáncer, disminuyendo de esta forma la tasa de detección de un lector determinado durante el periodo histórico.

El incremento del 16 % en la tasa de detección de cáncer descrito en este estudio no llega a tener significación estadística ya que la tasa de detección de cáncer histórica (5,25 ‰) está dentro de la horquilla del intervalo de confianza (5,1-7,2 ‰) obtenido en el estudio prospectivo. A pesar de ello, sí que puede considerarse relevante ya que implica que si todas las mamografías de la población histórica se hubiesen leído con el nuevo protocolo se habrían podido detectar hasta 40 cánceres más.

Diversos estudios de demostración inicial de beneficio potencial ^{112,113}, de valoración independiente del CAD ^{115,142,152} entre los que está el estudio 1 de esta tesis y otros estudios “de laboratorio” ^{119,120,121} han sido prometedores en cuanto a la habilidad del CAD para detectar cáncer de mama incluso de pequeño tamaño. Recientemente, Skaane *et al.* corroboraron que el CAD tenía el potencial para aumentar la detección de cáncer en un 16% y para reducir los cánceres de intervalo hasta en un 20 %. Sin embargo, reconocían que la

verdadera influencia e impacto de las marcas CAD en el proceso de decisión que lleva a cabo el radiólogo en cada mamografía solo podía ser valorado en un estudio prospectivo ¹⁵³.

El estudio 2 de esta tesis es uno de los pocos estudios prospectivos a gran escala realizados en cribado poblacional en Europa en los que se ha substituido una lectura por la asistencia del CAD y nuestros resultados no confirman el beneficio potencial que el CAD puede tener en la práctica clínica de un programa de cribado poblacional. Es decir, en el caso que el CAD hubiese marcado una serie de imágenes que pudieran corresponder a neoplasias, nuestros radiólogos expertos no consideraron esas marcas como valorables y por tanto tributarias de recitación.

Por tanto, en un contexto clínico real, teniendo como lectores radiólogos especializados en mama o radiólogos con buenos resultados en las auditorías anuales realizadas, el CAD no detectó más cánceres que el propio lector (tipo C de recitación: cero cánceres). Birdwell *et al.* en un contexto de hospital universitario concluyeron que el CAD aumentó la tasa de detección de cáncer en un 7,4 % aunque solo dos de sus 7 radiólogos estaban realmente especializados en imagen de la mama ¹²⁶. Por el contrario, 3 de los 4 radiólogos de nuestro grupo lector son o habían sido radiólogos de mama en una unidad asistencial por lo que habitualmente manejan recitaciones, realizan biopsias y tienen una retroalimentación continuada sobre la significación de las imágenes en las pacientes enviadas para ampliar estudio. Estos radiólogos asistenciales tenían una tasa de detección de cáncer mayor que los radiólogos generales y lo que es más importante, muestran una mayor confianza en su propio juicio, lo que hace que sean menos proclives a la influencia del CAD.

Muchos estudios han demostrado de forma fehaciente que el uso del CAD se acompaña de un aumento de la recitación. Gilbert *et al.* cuantificaron este incremento en un 15 % ¹⁵⁴ y

Cupples *et al.* en un 8,1 % ¹³⁴. Taylor y Potts, en una revisión sistemática de los estudios disponibles del efecto del CAD en la mamografía de cribado, concluyeron que el CAD aumenta la tasa de recitación sin tener un beneficio significativo en la detección ¹²⁷. Al contrario, Gromet afirmó que la recitación disminuía con la lectura simple con CAD lo que se relacionaba con que el brazo con doble lectura no disponía de arbitraje, lo que disparaba las recitaciones ¹³⁶. Nuestro estudio mostró un aumento de 3 puntos porcentuales en la recitación, que es más que lo publicado en otros estudios similares. No obstante, si solo tenemos en cuenta la tasa de recitación histórica de los lectores seleccionados antes del arbitraje, se observa que ha habido incluso una pequeña disminución en su tasa de recitación (de 7,24 % a 7,02 %) por lo que no creemos que el CAD haya contribuido a aumentar la tasa de recitación de estos radiólogos de “alto rendimiento” sino que, el arbitraje fue el responsable de la intensa reducción en la recitación durante el periodo histórico.

El programa de cribado poblacional en Barcelona incluye auditorías anuales como parte de los procesos de control de calidad, dirigidos entre otros, a mejorar el rendimiento de los radiólogos participantes. Los lectores de nuestro grupo reciben anualmente información de las neoplasias diagnosticadas y, en el periodo de doble lectura, de las lecturas falsas negativas, cuando las hubo. Además, los casos más interesantes, complejos o didácticos son discutidos en sesión para proporcionar formación continuada al grupo lector. Todos los radiólogos de nuestro grupo informaron una media de 2.700 mamografías de cribado anuales. Varios programas de cribado han establecido un mínimo de mamografías leídas para poder acreditarse como radiólogo lector en un programa de cribado de cáncer de mama entre 2.000 y 5.000 por año, aunque ninguno de ellos utilizó sus auditorías para seleccionar radiólogos ^{72,73}. Aunque los radiólogos de nuestro grupo cumplían el mínimo

establecido de 2.000 mamografías por año y recibieron información sobre neoplasias diagnosticadas, falsos negativos y formación continuada, solo aquellos con dedicación total a la imagen de la mama o muy motivados para mejorar su rendimiento cumplieron los criterios exigidos para entrar en el estudio de lectura única junto con CAD. Este resultado es consistente con lo reportado por Elmore *et al.* que observaron que los radiólogos que habían seguido un programa de especialización en mama (fellowship) tenían mejor sensibilidad, aunque la tasa de falsos positivos (más recitaciones) también era mayor ¹⁵⁵.

Rosenberg *et al.* recopilaron datos de 188 centros donde se realizaba mamografía de cribado incluyendo 807 radiólogos, entre 1996 y 2002, a través del Breast Cancer Surveillance Consortium (BCSC). Comunicaron una tasa media de detección de cáncer de 4,7 ‰ con una tasa de recitación de 9,8 % (6,4-13,3 %) ¹⁵⁶. Carney *et al.* encontraron una mediana (percentil 50) de 3,7 ‰ con una tasa de recitación del 9,2 % ¹⁵⁷. Estos dos estudios son de los pocos encontrados en la literatura dirigidos a investigar el rendimiento en la interpretación de mamografías de cribado y ambos se refieren a la realidad americana donde no existen verdaderos programas poblacionales de cribado, por lo que sus resultados no son directamente extrapolables. Por ello, para seleccionar el equipo de radiólogos lectores del estudio prospectivo, se utilizaron como referencia los resultados en nuestro medio. La tasa de detección de cáncer de mama en Cataluña en 2008-2009 fue de 3,91 ‰ (3,81 ‰ en segundo cribado o sucesivos) con una tasa de recitación de 4,9 % (4,4 % en segundo cribado o sucesivos) ¹⁵⁸. En la ciudad de Barcelona, la tasa de detección de cáncer de mama fue de 4,9 ‰ (4,4 ‰ en segundo cribado o sucesivos) ¹⁵⁹. A partir de estos datos se fijó el límite inferior que tenían que cumplir los radiólogos para ser seleccionados en el estudio en una tasa de detección de cáncer de 4,4 ‰ y una tasa de recitación entre 5-10 %. Aunque la guía europea recomienda una tasa de recitación inferior al 5 % en segundo

cribado o sucesivos ¹³⁸, se decidió fijar la nuestra en 5-10 % para priorizar la detección, ya que en general, los radiólogos con mayor tasa de detección de cáncer suelen tener también una mayor tasa de recitación. Es importante mencionar que los datos de rendimiento de los radiólogos de nuestro estudio en la fase histórica se ajustan a los descritos por Sickles *et al.*, que encontraron una tasa de detección de cáncer de 6 ‰ para radiólogos de mama y de 3,4 ‰ para radiólogos generales ¹⁶⁰.

La principal fortaleza de este estudio es que se diseñó de forma prospectiva, lo que permitió recoger de forma específica la aportación del CAD en cada recitación (recitaciones tipo A-B-C). Además, nuestro programa se adhiere a las recomendaciones de la guía europea por lo que se realizan auditorías anuales que permiten monitorizar de cerca el rendimiento del programa y de los propios radiólogos lectores lo que nos permitió hacer una selección de los mismos en base a su rendimiento previo histórico.

Una limitación de este estudio es que en Barcelona no existe un registro unificado de tumores para poder establecer la verdadera sensibilidad y especificidad tanto del programa como de los radiólogos. Por ello, se tomó la tasa de detección de cáncer y la tasa de recitación como estimadores indirectos de estas variables. Otro factor a comentar es el “factor laboratorio”. Como el comportamiento humano se modifica cuando un individuo se sabe observado, el comportamiento de los lectores también pudo ser diferente en este período en el que después de leer en parejas pasaron a leer solos apoyados por el sistema CAD. Así, aunque ninguno de los lectores admitió que el CAD hubiese modificado su intención de recitar en los casos que resultaron ser cáncer, sí es posible que de algún modo el CAD hubiese podido influir en ellos. Para clarificar este punto hubiese sido necesario que el sistema de informes hubiese permitido entrar dos informes diferentes (uno sin CAD y otro

con CAD) y que el CAD no hubiese estado disponible hasta que el radiólogo hubiese cerrado y validado su informe inicial.

VI- CONCLUSIONES

- 1- El CAD tiene una elevada sensibilidad y una muy baja especificidad para detectar cáncer de mama de pequeño tamaño en un programa de cribado poblacional. El elevado número de marcas falsas positivas constituye su principal limitación.

- 2- El CAD es muy fiable para señalar la presencia de grupos de calcificaciones y también aquellas lesiones tumorales que son hiperdensas en la mamografía. Por el contrario, su habilidad para detectar distorsiones es mucho menor.

- 3- Con el nuevo protocolo de lectura se observa un aumento en la detección de cáncer que es atribuible fundamentalmente a la selección de radiólogos más que al CAD propiamente dicho, que no detectó ningún cáncer que no hubiera sido previamente visto por el radiólogo. Por otro lado, la tasa de recitación aumentó de forma moderada, lo que probablemente se debe a la ausencia de árbitro, que tiene como papel fundamental disminuirla.

- 4- La detección de cáncer de mama con el nuevo protocolo se ha mantenido estable con tendencia a mejorar respecto a los valores históricos por lo que puede procederse a la implantación del nuevo protocolo de lectura sin merma en el número de cánceres diagnosticados siempre que se pueda realizar una adecuada selección de radiólogos y el área clínica pueda absorber un moderado aumento de la recitación.

VII- BIBLIOGRAFÍA

1. Departament de Salut. Generalitat de Catalunya. El càncer a Catalunya 1993-2020. [Internet]. 2012. Recuperado a partir de: <http://hdl.handle.net/11351/1056>
2. Sociedad Española de Oncología Médica (SEOM). Las Cifras del Cáncer en España 2014 [Internet]. 2014. Recuperado a partir de: http://www.seom.org/seomcms/images/stories/recursos/Las_cifras_del_cancer_2014.pdf
3. American Cancer Society. Breast Cancer [Internet]. 2013. Recuperado a partir de: <http://www.cancer.org/cancer/breastcancer/detailedguide/breast-cancer-pdf2>
4. American Cancer Society. Breast Cancer Facts and Figures 2013-2014 [Internet]. Atlanta. American Cancer Society, Inc. 2013. Recuperado a partir de: <http://www.cancer.org/Research/CancerFactsStatistics/ACSPC-042725>
5. National Cancer Institute. SEER Cancer Statistics Factsheets: Breast Cancer [Internet]. Recuperado a partir de: <http://seer.cancer.gov/statfacts/html/breast.html>
6. Tabar L, Duffy SW, Vitak B, Chen HH, Prevost TC. The natural history of breast carcinoma: what have we learned from screening? *Cancer*. 1999;86(3):449-62.
7. Duffy SW, Tabar L, Vitak B, Day NE, Smith RA, Chen HH, et al. The relative contributions of screen-detected in situ and invasive breast carcinomas in reducing mortality from the disease. *Eur J Cancer*. 2003;39(12):1755-60.
8. Chan DSM, Vieira a R, Aune D, Bandera E V, Greenwood DC, McTiernan A, et al. Body mass index and survival in women with breast cancer-systematic literature review and meta-analysis of 82 follow-up studies. *Ann Oncol*. 2014;10:1901-14.
9. Chlebowski RT, Blackburn GL, Thomson CA, Nixon DW, Shapiro A, Hoy MK, et al. Dietary fat reduction and breast cancer outcome: Interim efficacy results from the women's intervention nutrition study. *J Natl Cancer Inst*. 2006;98(24):1767-76.
10. Ballard-Barbash R, Hunsberger S, Alciati MH, Blair SN, Goodwin PJ, McTiernan A, et al. Physical activity, weight control, and breast cancer risk and survival: Clinical trial rationale and design considerations. *J Natl Cancer Inst*. 2009;101(9):630-43.
11. Galceran J, Ameijide A, Cardó X, Piñol JL, Gumà J SF et al. El càncer a Tarragona, 1980-2001. Incidència, mortalitat, supervivència i prevalença. *Regist Càncer Tarragona Fund lliga per a la Investigació i prevenció del càncer*, Reus. 2008.
12. Frame PS, Carlson SJ. A critical review of periodic health screening using specific screening criteria. Part 4: selected miscellaneous diseases. *J Fam Pract*. 1975;2(4):283-9.
13. Shapiro S, Venet W, Strax P, Venet L, Roeser R. Ten- to fourteen-year effect of screening on breast cancer mortality. *J Natl Cancer Inst*. 1982;69(2):349-55.
14. Tabár L, Gad a, Holmberg L., Ljungquist U, Fagerberg CJ., Baldetorp L, et al. Reduction in Mortality From Breast Cancer After Mass Screening With Mammography. *Lancet*. 1985;1(8433):829-32.

15. Nystrom L, Rutqvist LE, Wall S, Lindgren A, Lindqvist M, Ryden S, et al. Breast cancer screening with mammography: Overview of Swedish randomised trials. *Lancet*. 1993;341(8851):973-8.
16. Larsson LG, Nyström L, Wall S, Rutqvist L, Andersson I, Bjurstam N, et al. The Swedish randomised mammography screening trials: analysis of their effect on the breast cancer related excess mortality. *J Med Screen*. 1996;3(3):129-32.
17. Miller AB, To T, Baines CJ, Wall C. Canadian National Breast Screening Study-2: 13-year results of a randomized trial in women aged 50-59 years. *J Natl Cancer Inst*. 2000;92(18):1490-9.
18. Roberts MM, Alexander FE, Anderson TJ, Chetty U, Donnan PT, Forrest P, et al. Edinburgh trial of screening for breast cancer: mortality at seven years. *Lancet*. 1990;335(8684):241-6.
19. Smith R a., Duffy SW, Gabe R, Tabar L, Yen AMF, Chen THH. The randomized trials of breast cancer screening: What have we learned? *Radiol Clin North Am*. 2004;42(5):793-806.
20. Kopans DB, Feig S a. The Canadian National Breast Screening Study: a critical review. *AJR Am J Roentgenol*. 1993;161(4):755-60.
21. Gotzsche P and O. Is screening for breast cancer with mammography really justifiable? *Lancet*. 2000;355(9198):129-34.
22. Olsen O, Gøtzsche PC. Cochrane review on screening for breast cancer with mammography. *Lancet*. 2001;358(9290):1340-2.
23. Duffy SW, Tabar L and SR. The mammographic Screening Trials: Commentary on the Recent Work by Olsen and Gotzsche. *J Surg Oncol*. 2002;81(4):159.
24. Tabar L, Vitak B, Chen HH, Yen MF, Duffy SW, Smith RA. Beyond randomized controlled trials: organized mammographic screening substantially reduces breast carcinoma mortality. *Cancer*. 2001;91(9):1724-31.
25. Nyström L, Andersson I, Bjurstam N, Frisell J, Nordenskjöld B, Rutqvist LE. Long-term effects of mammography screening: Updated overview of the Swedish randomised trials. *Lancet*. 2002;359(9310):909-19.
26. Vainio H, Bianchini F E. Breast cancer screening. IARC Handbooks of Cancer Prevention, volume 7. Lyon: IARC press; 2002.
27. The canadian task force on preventive health care. Recommendations on screening for breast cancer in average-risk women aged 40-74 years. *Can Med Assoc J*. 2011;183:1991-2001.
28. Nelson HD, Tyne K, Naik A, Bougatsos C, Chan BK, Humphrey L. Screening for breast cancer: An update for the U.S. Preventive Services Task Force. *Ann Intern Med*. 2009;151(10):727-37.
29. Duffy SW, Tabár L, Chen HH, Holmqvist M, Yen MF, Abdsalah S, et al. The impact of organized mammography service screening on breast carcinoma mortality in seven Swedish Counties: A collaborative evaluation. *Cancer*. 2002;95(3):458-69.

30. World Health Organization. WHO Position Paper on Mammography Screening [Internet]. 2014. Recuperado a partir de:
http://apps.who.int/iris/bitstream/10665/137339/1/9789241507936_eng.pdf?ua=1
31. Independent UK Panel on Breast Cancer Screening. The benefits and harms of breast cancer screening: an independent review. *Lancet*. 2012;380(9855):1778-86.
32. Kopans DB. *La mama en imagen*. 1999. Ed. Marban Libros S.L.; 1999. 55-100 p.
33. Webb M, Cady B, Michaelson J. A failure analysis of invasive breast cancer. *Cancer*. 2013;120(18):2839-46.
34. Moss SM, Nystrom L, Jonsson H, Paci E, Lynge E, Njor S, et al. The impact of mammographic screening on breast cancer mortality in Europe: a review of trend studies. *J Med Screen*. 2012;19 (Supplement 1):26-32.
35. Biesheuvel C, Barratt A, Howard K, Houssami N, Irwig L. Effects of study methods and biases on estimates of invasive breast cancer overdiagnosis with mammography screening: a systematic review. *Lancet Oncol*. 2007;8(12):1129-38.
36. Zackrisson S. Rate of over-diagnosis of breast cancer 15 years after end of Malmo mammographic screening trial: follow-up study. *BMJ*. 2006;332 (7543):689-92.
37. Duffy SW, Tabar L, Olsen AH, Vitak B, Allgood PC, Chen THH, et al. Absolute numbers of lives saved and overdiagnosis in breast cancer screening, from a randomized trial and from the Breast Screening Programme in England. *J Med Screen*. 2010;17(1):25-30.
38. Jørgensen KJ, Gøtzsche PC. Overdiagnosis in publicly organised mammography screening programmes: systematic review of incidence trends. *BMJ*. 2009;339:b2587.
39. Bleyer A, Welch HG. Effect of three decades of screening mammography on breast-cancer incidence. *N Engl J Med*. 2012;367(21):1998-2005.
40. Helvie MA, Chang JT, Hendrick RE, Banerjee M. Reduction in late-stage breast cancer incidence in the mammography era: Implications for overdiagnosis of invasive cancer. *Cancer*. 2014;120(17):2649-56.
41. Olsen AH, Agbaje OF, Myles JP, Lynge E, Duffy SW. Overdiagnosis, sojourn time, and sensitivity in the Copenhagen mammography screening program. *Breast J*. 2006;12(4):338-42.
42. Puliti D, Miccinesi G, Zappa M, Manneschi G, Crocetti E, Paci E. Balancing harms and benefits of service mammography screening programs: a cohort study. *Breast Cancer Res*.; 2012;14(1):R9.
43. Page DL, Dupont WD, Rogers LW, Landenberger M. Intraductal carcinoma of the breast: follow-up after biopsy only. *Cancer*. 1982;49(4):751-8.
44. Sanders ME, Schuyler PA, Dupont WD, Page DL. The natural history of low-grade ductal carcinoma in situ of the breast in women treated by biopsy only revealed over 30 years of long-term follow-up. *Cancer*. 2005;103(12):2481-4.

45. Cuzick J, Sestak I, Pinder SE, Ellis IO, Forsyth S, Bundred NJ, et al. Effect of tamoxifen and radiotherapy in women with locally excised ductal carcinoma in situ: Long-term results from the UK/ANZ DCIS trial. *Lancet Oncol.* 2011;12(1):21-9.
46. Puliti D, Miccinesi G, Paci E. Overdiagnosis in breast cancer: Design and methods of estimation in observational studies. *Prev Med (Baltim).* 2011;53(3):131-3.
47. Hofvind S, Ponti a., Patnick J, Ascunce N, Njor S, Broeders M, et al. False-positive results in mammographic screening for breast cancer in Europe: a literature review and survey of service screening programmes. *J Med Screen.* 2012;19(Supplement 1):57-66.
48. Hofvind S, Skaane P, Vitak B, Wang H, Thoresen S, Eriksen L, et al. Influence of review design on percentages of missed interval breast cancers: retrospective study of interval cancers in a population-based screening program. *Radiology.* 2005;237(2):437-43.
49. Törnberg S, Kemetli L, Ascunce N, Hofvind S, Anttila A, Sèradour B, et al. A pooled analysis of interval cancer rates in six European countries. *Eur J Cancer Prev.* 2010;19(2):87-93.
50. Departament de Salut. Generalitat de Catalunya. El cribratge de càncer de mama: revisió dels beneficis i efectes adversos [Internet]. 2014. Recuperado a partir de: http://cancer.gencat.cat/web/sites/cancer/ca/ciutadans/deteccio_precoc/programes_de_deteccio_precoc/cancer_de_mama/Eficacia-i-efectivitat-del-cribratge-cancer-mama-2014-corregit-DS-2_2.pdf
51. American College of Radiology. Breast Imaging Reporting and Data System (BIRADS). 4th Ed. Va; 2003.
52. Butlletí Informatiu de l'Oficina Tècnica del Programa de Detecció Precoç de Càncer de Mama de Barcelona. Barcelona. Num.10 [Internet]. Agència de Salut Pública de Barcelona. 2014. Recuperado a partir de: http://www.aspb.cat/quefem/docs/butlleti_10_cancer_mama.pdf
53. Hendrick RE. Radiation doses and cancer risks from breast imaging studies. *Radiology.* 2010;257(1):246-53.
54. American Cancer Society. Mammograms and Other Breast Imaging Procedures. [Internet]. 2014. Recuperado a partir de: <http://www.cancer.org/healthy/findcancerearly/examandtestdescriptions/mammogramsandotherbreastimagingprocedures/mammograms-and-other-breast-imaging-procedures-pdf>
55. Dershaw DD. Status of mammography after the Digital Mammography Imaging Screening Trial: Digital versus film. *Breast J.* 2006;12(2):99-102.
56. Karssemeijer N, Frieling JT, Hendriks JH. Spatial resolution in digital mammography. *Invest Radiol.* 1993;28(5):413-9.
57. Obenauer S, Luftner-Nagel S, von Heyden D, Munzel U, Baum E, Grabbe E. Screen film vs full-field digital mammography: Image quality, detectability and characterization of lesions. *Eur Radiol.* 2002;12(7):1697-702.

58. Fischmann A, Siegmann KC, Wersebe A, Claussen CD, Müller-Schimpfle M. Comparison of full-field digital mammography and film-screen mammography: Image quality and lesion detection. *Br J Radiol.* 2005;78(928):312-5.
59. Chevalier M, Torres R. Mamografía digital. *Rev Fis Med.* 2010;11(1):11-26.
60. Chiarelli AM, Edwards S a, Prummel M V, Muradali D, Majpruz V, Done SJ, et al. Digital compared with screen-film mammography: performance measures in concurrent cohorts within an organized breast screening program. *Radiology.* 2013;268(3):684-93.
61. Torres Cabrera R. Mamografía digital: estado actual de la tecnología. *Radiologia.* 2010;52(S1):7-13.
62. Skaane P, Young K, Skjennald A. Population-based mammography screening: comparison of screen-film and full-field digital mammography with soft-copy reading--Oslo I study. *Radiology.* 2003;229(3):877-84.
63. P Skaane and Skjennald A. Screen-film mammography versus full-field digital mammography with soft-copy reading: Randomized trial in a population-based screening program - The Oslo II study. *Radiology.* 2004;232:197-204.
64. Pisano ED, Gatsonis C, Hendrick E, Yaffe M, Baum JK, Acharyya S, et al. Diagnostic performance of digital versus film mammography for breast-cancer screening. *N Engl J Med.* 2005;353(17):1773-83.
65. Holland R, Hendriks JH, Mravunac M. Mammographically occult breast cancer. A pathologic and radiologic study. *Cancer.* 1983;52(10):1810-9.
66. Kolb TM, Lichy J, Newhouse JH. Comparison of the performance of screening mammography, physical examination, and breast US and evaluation of factors that influence them: an analysis of 27,825 patient evaluations. *Radiology.* 2002;225(1):165-75.
67. Mandelson MT, Oestreicher N, Porter PL, White D, Finder CA, Taplin SH, et al. Breast density as a predictor of mammographic detection: comparison of interval- and screen-detected cancers. *J Natl Cancer Inst.* 2000;92(13):1081-7.
68. Carney PA, Miglioretti DL, Yankaskas BC, Kerlikowske K, Rosenberg R, Rutter CM, et al. Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography. *Ann Intern Med.* 2003;138(3):168-75.
69. Buist DSM, Porter PL, Lehman C, Taplin SH, White E. Factors contributing to mammography failure in women aged 40-49 years. *J Natl Cancer Inst.* 2004;96(19):1432-40.
70. Sankatsing VD, Heijnsdijk E a., van Luijt P a., van Ravesteyn NT, Fracheboud J, de Koning HJ. Cost-effectiveness of digital mammography screening before the age of 50 in the Netherlands. *Int J Cancer.* 2015;137(8):1990-9.
71. Bargallo, X, Santamaria, G, Cores, A, del Amo, M, Arguis, P, Velasco M. False Negative Readings of Breast Cancer. En: *Radiological Society of North America 2007 Scientific Assembly and Annual Meeting.* Chicago IL. 2007.

72. Cornford E, Reed J, Murphy A, Bennett R, Evans A. Optimal screening mammography reading volumes; evidence from real life in the East Midlands region of the NHS Breast Screening Programme. *Clin Radiol*. 2011;66:103-7.
73. Rawashdeh M, Lee WB, Bourne RM, Ryan EA, Reed WM, Heard RC, et al. Markers of good performance in mammography depend on number of annual readings. *Radiology*. 2013;269(1):61-7.
74. Beam CA, Conant EF, Sickles EA. Association of volume and volume-independent factors with accuracy in screening mammogram interpretation. *J Natl Cancer Inst*. 2003;95(4):282-90.
75. Baker JA, Rosen EL, Lo JY, Gimenez EI, Walsh R, Soo MS. Computer-aided detection (CAD) in screening mammography: sensitivity of commercial CAD systems for detecting architectural distortion. *AJR Am J Roentgenol*. 2003;181(4):1083-8.
76. Vitak B. Invasive interval cancers in the Ostergötland Mammographic Screening Programme: radiological analysis. *Eur Radiol*. 1998;8(4):639-46.
77. Yankaskas BC, Schell MJ, Bird RE, Desrochers DA. Reassessment of breast cancers missed during routine screening mammography: A community-based study. *AJR Am J Roentgenol*. 2001;177(3):535-41.
78. Birdwell RL, Ikeda DM, O'Shaughnessy KF, Sickles E a. Mammographic characteristics of 115 missed cancers later detected with screening mammography and the potential utility of computer-aided detection. *Radiology*. 2001;219(1):192-202.
79. Kemp Jacobsen K, O'Meara ES, Key D, S.M. Buist D, Kerlikowske K, Vejborg I, et al. Comparing sensitivity and specificity of screening mammography in the United States and Denmark. *Int J Cancer*. 2015;137(9):2198-207.
80. Beam CA, Layde PM, Sullivan DC. Variability in the interpretation of screening mammograms by US radiologists. Findings from a national sample. *Arch Intern Med*. 1996;156(2):209-13.
81. Kopans DB. Double reading. *Radiol Clin North Am*. 2000;38(4):719-24.
82. Thurffjell EL, Lernevall KA, Taube AA. Benefit of independent double reading in a population-based mammography screening program. *Radiology*. 1994;191:241-4.
83. Anttinen I, Pamilo M, Soiva M, Roiha M. Double reading of mammography screening films--one radiologist or two? *Clin Radiol*. 1993;48(6):414-21.
84. Warren RML, Duffy SW. Comparison of single reading with double reading of mammograms, and change in effectiveness with experience. *Br J Radiol*. 1995;68(813):958-62.
85. Taplin SH, Rutter CM, Elmore JG, Seger D, White D, Brenner RJ. Accuracy of screening mammography using single versus independent double interpretation. *AJR Am J Roentgenol*. 2000;174(5):1257-62.
86. Mucci B, Athey G, Scarisbrick G. Double read screening mammograms: The use of a third reader to arbitrate on disagreements. *Breast*. 1999;8(1):37-9.

87. Duijm LEM, Louwman MWJ, Groenewoud JH, van de Poll-Franse L V, Fracheboud J, Coebergh JW. Inter-observer variability in mammography screening and effect of type and number of readers on screening outcome. *Br J Cancer*. 2009;100:901-7.
88. Bennett RL, Sellars SJ, Blanks RG, Moss SM. An observational study to evaluate the performance of units using two radiographers to read screening mammograms. *Clin Radiol*. 2012;67(2):114-21.
89. D'Orsi C, Tu S-P, Nakano C, Carney P a, Abraham L a, Taplin SH, et al. Current realities of delivering mammography services in the community: do challenges with staffing and scheduling exist? *Radiology*. 2005;235(2):391-5.
90. Duijm LEM, Groenewoud JH, Fracheboud J, de Koning HJ. Additional double reading of screening mammograms by radiologic technologists: Impact on screening performance parameters. *J Natl Cancer Inst*. 2007;99(15):1162-70.
91. Tonita JM, Hillis JP, Lim CH. Medical radiologic technologist review: effects on a population-based breast cancer screening program. *Radiology*. 1999;211(2):529-33.
92. Wivell G, Denton ERE, Eve CB, Inglis JC, Harvey I. Can radiographers read screening mammograms? *Clin Radiol*. 2003;58(1):63-7.
93. Liston JC, Dall BJG. Can the NHS Breast Screening Programme afford not to double read screening mammograms? *Clin Radiol*. 2003;58(6):474-7.
94. Ciatto S, Ambrogetti D, Risso G, Catarzi S, Morrone D, Mantellini P, et al. The role of arbitration of discordant reports at double reading of screening mammograms. *J Med Screen*. 2005;12(3):125-7.
95. Duijm LEM, Groenewoud JH, Hendriks JHCL, de Koning HJ. Independent double reading of screening mammograms in The Netherlands: effect of arbitration following reader disagreements. *Radiology*. 2004;231(2):564-70.
96. Gur D, Stalder JS, Hardesty LA, Zheng B, Sumkin JH, Chough DM, et al. Computer-aided detection performance in mammographic examination of masses: assessment. *Radiology*. 2004;233(2):418-23.
97. Klompenhouwer EG, Duijm LEM, Voogd a. C, den Heeten GJ, Nederend J, Jansen FH, et al. Variations in screening outcome among pairs of screening radiologists at non-blinded double reading of screening mammograms: a population-based study. *Eur Radiol*. 2014;24(5):1097-104.
98. Beam CA, Sullivan DC, Layde PM. Effect of human variability on independent double reading in screening mammography. *Acad Radiol*. 1996;3(11):891-7.
99. F Winsberg, M Elkin, J Macy VB and WW. Detection of radiographic abnormalities in mammograms by means of optical scanning and computer analysis. *Radiology*. 1967;89:211-5.
100. Chan H, Doi K, Galhotra S, Viborny C, MacMahon H, Jokich P. Image feature analysis and computer-aided diagnosis in digital radiography. 1. Automated detection of microcalcifications in mammography. *Med Phys*. 1987;14:538-48.

101. Chan HP, Doi K, Vyborny CJ, Lam KL, Schmidt RA. Computer-aided detection of microcalcifications in mammograms. Methodology and preliminary clinical study. *Invest Radiol*. 1988;23(9):664-71.
102. Doi K. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Comput Med Imaging Graph*. 2007;31(4-5):198-211.
103. Nishikawa RM, Doi K, Giger ML, Schmidt RA, Vyborny CJ, Monnier-Cholley L, et al. Computerized detection of clustered microcalcifications: evaluation of performance on mammograms from multiple centers. *Radiographics*. 1995;15(2):443-52.
104. Giger ML, Chan H-P, Boone J. Anniversary paper: History and status of CAD and quantitative image analysis: the role of Medical Physics and AAPM. *Med Phys*. 2008;35(12):5799-820.
105. Oliver A, Freixenet J, Martí J, Pérez E, Pont J, Denton ERE, et al. A review of automatic mass detection and segmentation in mammographic images. *Med Image Anal*. 2010;14(2):87-110.
106. Bargalló X, Santamaría G, Velasco M, del Amo M, Arguis P, Burrel M, et al. Mammographic features of screening detected pT1 (a-b) invasive breast cancer using BI-RADS lexicon. *Eur J Radiol*. 2012;81(10):2620-6.
107. Tang J, Rangayyan RM, Xu J, El Naqa I, Yang Y. Computer-aided detection and diagnosis of breast cancer with mammography: recent advances. *IEEE Trans Inf Technol Biomed*. 2009;13(2):236-51.
108. Vyborny CJ, Giger ML. Computer vision and artificial intelligence in mammography. *AJR Am J Roentgenol*. 1994;162(3):699-708.
109. Rangayyan RM, Ayres FJ, Desautels JEL. A review of computer-aided diagnosis of breast cancer: toward the detection of subtle signs. *J Franklin Institute-Engineering Appl Math*. 2007;344(3-4):312-48.
110. Zheng B. Mass detection scheme for digitized mammography. Hayat H, editor. *Cancer Imaging: lung and breast carcinomas, Vol 1*. Hayat, HA. Elsevier Academic Press; 2008. 325-337 p.
111. Wei J, Chan HP, Zhou C, Wu YT, Sahiner B, Hadjiiski LM, et al. Computer-aided detection of breast masses: four-view strategy for screening mammography. *Med Phys*. 2011;38(4):1867-76.
112. Warren Burhenne LJ, Wood SA, D'Orsi CJ, Feig SA, Kopans DB, O'Shaughnessy KF, et al. Potential contribution of computer-aided detection to the sensitivity of screening mammography. *Radiology*. 2000;215:554-62.
113. Brem RF, Baum J, Lechner M, Kaplan S, Souders S, Naul LG, et al. Improvement in sensitivity of screening mammography with computer-aided detection: A multiinstitutional trial. *AJR Am J Roentgenol*. 2003;181(3):687-93.
114. Houssami N, Given-Wilson R. Incorporating new technologies into clinical practice without evidence of effectiveness in prospective studies: computer-aided detection (CAD) in breast screening reinforces the need for better initial evaluation. *Breast*. 2007;16(3):219-21.

115. Ellis RL, Meade AA, Mathiason MA, Willison KM, Logan-Young W. Evaluation of computer-aided detection systems in the detection of small invasive breast carcinoma. *Radiology*. 2007;245(1):88-94.
116. Kim SJ, Moon WK, Kim SY, Chang JM, Kim SM, Cho N. Comparison of two software versions of a commercially available computer-aided detection (CAD) system for detecting breast cancer. *Acta Radiol*. 2010;51(5):482-90.
117. Soo MS, Rosen EL, Xia JQ, Ghate S, Baker JA. Computer-aided detection of amorphous calcifications. *AJR Am J Roentgenol*. 2005;184(3):887-92.
118. Hupse R, Samulski M, Lobbes M, Den Heeten A, Imhof-Tas MW, Beijerinck D, et al. Standalone computer-aided detection compared to radiologists' performance for the detection of mammographic masses. *Eur Radiol*. 2013;23(1):93-100.
119. Chan HP, Doi K, Vyborny CJ, Schmidt RA, Metz CE, Lam KL, et al. Improvement in radiologists' detection of clustered microcalcifications on mammograms. The potential of computer-aided diagnosis. *Invest Radiol*. 1990;25(10):1102-10.
120. Jiang Y, Nishikawa RM, Schmidt RA, Metz CE, Giger ML, Doi K. Improving breast cancer diagnosis with computer-aided diagnosis. *Acad Radiol*. 1999;6(1):22-33.
121. Huo Z, Giger ML, Vyborny CJ, Metz CE. Breast cancer: effectiveness of computer-aided diagnosis observer study with independent database of mammograms. *Radiology*. 2002;224(2):560-8.
122. Houssami N, Given-Wilson R, Ciatto S. Early detection of breast cancer: overview of the evidence on computer-aided detection in mammography screening. *J Med Imaging Radiat Oncol*. 2009;53(2):171-6.
123. Taylor P, Champness J, Given-Wilson R, Johnston K, Potts H. Impact of computer-aided detection prompts on the sensitivity and specificity of screening mammography. *Health Technol Assess*. 2005;9(6):III, 1-58.
124. Freer TW, Ulissey MJ. Screening mammography with computer-aided detection: prospective study of 12,860 patients in a community breast center. *Radiology*. 2001;220(3):781-6.
125. Morton MJ, Whaley DH, Brandt KR, Amrami KK. Screening mammograms: interpretation with computer-aided detection--prospective evaluation. *Radiology*. 2006;239(2):375-83.
126. Birdwell RL, Bandodkar P, Ikeda DM. Computer-aided detection with screening mammography in a university hospital setting. *Radiology*. 2005;236(2):451-7.
127. Taylor P, Potts HW. Computer aids and human second reading as interventions in screening mammography: two systematic reviews to compare effects on cancer detection and recall rate. *Eur J Cancer*. 2008;44(6):798-807.
128. Noble M, Bruening W, Uhl S, Schoelles K. Computer-aided detection mammography for breast cancer screening: systematic review and meta-analysis. *Arch Gynecol Obstet*. 2009;279:881-90.

129. Fenton JJ, Taplin SH, Carney PA, Abraham L, Sickles EA, D'Orsi C, et al. Influence of computer-aided detection on performance of screening mammography. *N Engl J Med.* 2007;356:1399-409.
130. Khoo LA, Taylor P, Given-Wilson RM. Computer-aided detection in the United Kingdom National Breast Screening Programme: prospective study. *Radiology.* 2005;237(2):444-9.
131. Gilbert FJ, Astley SM, McGee MA, Gillan MG, Boggis CR, Griffiths PM, et al. Single reading with computer-aided detection and double reading of screening mammograms in the United Kingdom National Breast Screening Program. *Radiology.* 2006;241(1):47-53.
132. Georgian-Smith D, Moore RH, Halpern E, Yeh ED, Rafferty E a., D'Alessandro HA, et al. Blinded comparison of computer-aided detection with human second reading in screening mammography. *AJR Am J Roentgenol.* 2007;189:1135-41.
133. Gilbert FJ, Astley SM, Gillan MGC, Agbaje OF, Wallis MG, James J, et al. Single reading with computer-aided detection for screening mammography. *N Engl J Med.* 2008;359:1675-84.
134. Cupples TE, Cunningham JE, Reynolds JC. Impact of computer-aided detection in a regional screening mammography program. *AJR Am J Roentgenol.* 2005;185(4):944-50.
135. Gur D, Sumkin JH, Rockette HE, Ganott M, Hakim C, Hardesty L, et al. Changes in Breast Cancer Detection and Mammography Recall Rates After the Introduction of a Computer-Aided Detection System. *J Natl Cancer Inst.* 2004;96(3):185-90.
136. Gromet M. Comparison of computer-aided detection to double reading of screening mammograms: review of 231,221 mammograms. *AJR Am J Roentgenol.* 2008;190(4):854-9.
137. Azavedo E, Zackrisson S, Mejàre I, Heibert Arnlin M. Is single reading with computer-aided detection (CAD) as good as double reading in mammography screening? A systematic review. *BMC Med Imaging.* 2012;12:22.
138. Perry N de Wolf C, et al, eds. BM. European guidelines for quality assurance in mammography screening. Fourth edition. Luxembourg: Office for Official Publications of the European Communities; 2006.
139. Birdwell RL. The preponderance of evidence supports computer-aided detection for screening mammography. *Radiology* 2009;253(1):9-16.
140. Yang SK, Moon WK, Cho N, Park JS, Cha JH, Kim SM, et al. Screening mammography-detected cancers: sensitivity of a computer-aided detection system applied to full-field digital mammograms. *Radiology.* 2007;244:104-11.
141. The JS, Schilling KJ, Hoffmeister JW, Friedmann E, McGinnis R, Holcomb RG. Detection of breast cancer with full-field digital mammography and computer-aided detection. *AJR Am J Roentgenol.* 2009;192(2):337-40.
142. Sadaf A, Crystal P, Scaranelo A, Helbich T. Performance of computer-aided detection applied to full-field digital mammography in detection of breast cancers. *Eur J Radiol.* 2011;77(3):457-61.

143. Weigel S, Decker T, Korsching E, Hungermann D, Bocker W, Heindel W. Calcifications in digital mammographic screening: improvement of early detection of invasive breast cancers? *Radiology*. 2010;255(3):738-45.
144. Bagnall MJ, Evans AJ, Wilson AR, Pinder SE, Denley H, Geraghty JG, et al. Predicting invasion in mammographically detected microcalcification. *Clin Radiol*. 2001;56(10):828-32.
145. Stomper PC, Connolly JL. Mammographic features predicting an extensive intraductal component in early-stage infiltrating ductal carcinoma. *AJR Am J Roentgenol*. 1992;158(2):269-72.
146. Bolivar A V, Gomez SS, Merino P, Alonso-Bartolom P, Garcia EO, Cacho PM, et al. Computer-aided detection system applied to full-field digital mammograms. *Acta Radiol*. 2010;51(10):1086-92.
147. Malich A, Fischer DR, Bottcher J. CAD for mammography: the technique, results, current role and further developments. *Eur Radiol*. 2006;16(7):1449-60.
148. Brem RF, Hoffmeister JW, Rapelyea JA, Zisman G, Mohtashemi K, Jindal G, et al. Impact of breast density on computer-aided detection for breast cancer. *AJR Am J Roentgenol*. 2005;184(2):439-44.
149. Ho WT, Lam PW. Clinical performance of computer-assisted detection (CAD system in detecting carcinoma in breasts of different densities. *Clin Radiol*. 2003;58(2):133-6.
150. Philpotts LE. Can computer-aided detection be detrimental to mammographic interpretation? *Radiology*. 2009;253(1):17-22.
151. Taylor P, Given-Wilson R, Champness J, Potts HW, Johnston K. Assessing the impact of CAD on the sensitivity and specificity of film readers. *Clin Radiol*. 2004;59(12):1099-105.
152. Kim SJ, Moon WK, Cho N, Cha JH, Kim SM, Im JG. Computer-aided detection in digital mammography: comparison of craniocaudal, mediolateral oblique, and mediolateral views. *Radiology*. 2006;241(3):695-701.
153. Skaane P, Kshirsagar A, Hofvind S, Jahr G, Castellino RA. Mammography screening using independent double reading with consensus: Is there a potential benefit for computer-aided detection? *Acta radiol*. 2012;53(3):241-8.
154. Gilbert FJ, Astley SM, McGee MA, Gillan MGC, Boggis CRM, Griffiths PM, et al. Single reading with computer-aided detection and double reading of screening mammograms in the United Kingdom National Breast Screening Program. *Radiology*. 2006;241(1):47-53.
155. Elmore JG, Jackson SL, Abraham L, Miglioretti DL, Carney PA, Geller BM, et al. Variability in interpretive performance at screening mammography and associated with accuracy. *Radiology*. 2009;253(3):641-51.
156. Rosenberg RD, Yankaskas BC, Abraham LA, Sickles EA, Lehman CD, Geller BM, et al. Performance benchmarks for screening mammography. *Radiology*. 2006;241(1):55-66.

157. Carney PA, Sickles EA, Monsees BS, Bassett LW, Brenner RJ, Feig SA, et al. Identifying minimally acceptable interpretive performance criteria for screening mammography. *Radiology*. 2010;255(2):354-61.
158. Asuncion N, Delgado J, Salas D, Zubizarreta R, Ederri M. Programas de detección precoz de cáncer de mama en España: características y principales resultados. *Med Clin (Barc)*. 2013;141(1):13-23.
159. Butlletí Informatiu de l'Oficina Tècnica del Programa de Detecció Precoç de Càncer de Mama de Barcelona. Barcelona. Num.6 [Internet]. Agència de Salut Pública de Barcelona. 2011. Recuperado a partir de: http://www.aspb.cat/quefem/docs/butlleti_6_cancer_mama.pdf
160. Sickles EA, Wolverton DE, Dee KE. Performance parameters for screening and diagnostic mammography: specialist and general radiologists. *Radiology*. 2002;224:861-9.